

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 10, 2005, 05:44:18 : Search time 475 Seconds
(without alignments)
9346.960 Million cell updates/sec

Title: US-10-620-487-1

Perfect score: 750
Sequence: 1 gaccataaccaccaacta.....ttagtaactgctacttaag 750

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq.16Dec04:*
1: geneseqn1980s:*
2: geneseqn1980s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001as:*
6: geneseqn2002as:*
7: geneseqn2002as:*
8: geneseqn2003as:*
9: geneseqn2003as:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	750	100.0	750	12	ADK66005
2	588	78.4	588	3	AA552690
3	295.8	39.4	1668	5	AA581979
4	265.8	35.4	618	11	ACH99695
5	156	20.8	372	11	ACH99702
6	100	13.3	100	8	ACD79726
7	100	13.3	100	8	ACD79725
8	100	13.3	100	8	ACD79727
9	50.4	6.7	573	6	ABK73143
10	42.6	5.7	2325	5	AA592926
11	42.6	5.7	3020	5	AA588917
12	42	5.6	669	9	ADA31058
13	39.8	5.3	636	10	ADP02251
14	39.6	5.3	750	11	ABD10542
15	39	5.2	247682	12	ADL08109
16	38.2	5.1	636	11	ACH97476
17	37.2	5.0	609	10	ACF70009
18	37.2	5.0	110000	10	ACF65385_1
19	37.2	5.0	110000	10	ACF67367_30
20	36.4	4.9	2000	8	ADA71938

ALIGNMENTS

RESULT 1	ID	ADK66005	ADK66005 standard; DNA; 750 BP.	AD62125	Ad62125 Cotton cD
XX	AC	ADK66005;		492	ACD96860
XX	DT	06-MAY-2004 (first entry)		612	ACF71507
XX	DE	E coli yfik gene.		110000	ACF67367_46
XX	KW	ds; gene; yfik; phosphoglycerate amino acid; amino acid production.		4.7	110000
XX	OS	Escherichia coli.		4.7	110000
XX	FT	Key	Location/Qualifiers	4.7	110000
XX	FT	CDS	110..697	4.7	110000
XX	FT		/*tag= a	4.7	110000
XX	FT		/product= "yfiK"	4.7	110000
XX	PN	EP1382684-A1.		4.7	110000
XX	PD	21-JAN-2004.		4.7	110000
XX	PF	10-JUL-2003; 2003BP-00015546.		4.7	110000
XX	PR	19-JUL-2002; 2002DE-01032930.		4.7	110000
XX	PA	(CONE) CONSORTIUM ELEKTROCHEM IND GMBH.		4.7	110000
XX	PI	Water T;		4.7	110000
XX	DR	WPI; 2004-192974/19.		4.7	110000
XX	DR	P-PSDB; ADK66006.		4.7	110000
XX	PT	New microbial strain, useful for producing amino acids of the		4.7	110000
XX	PT	phosphoglycerate family, particularly serine and cysteine, has increased		4.7	110000
XX	PS	activity of the yfik gene product.		4.7	110000
XX	PS	Disclosure; Page 8-10; 16pp; German.		4.7	110000
XX	CC	The present invention relates to a microbial strain for fermentative		4.7	110000
XX	CC	product of amino acids of the phosphoglycerate family, or their		4.7	110000
XX	CC	derivatives, has higher activity of the product of the yfik gene, or its		4.7	110000

CC homologue, compared with its parent strain. The microbial strain is used
 CC to produce O- and N-acetyl-serine, L-cysteine and cystine. The present
 CC sequence is the E. coli yfik gene.

XX
 SQ Sequence 750 BP; 172 A; 169 C; 194 G; 215 T; 0 U; 0 Other;

Query Match 100.0%; Score 750; DB 12; Length 750;

Best Local Similarity 100.0%; Pred. No. 1.5e-247;
 Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GATCATTAACCCCAACCTATCGAAATATGATCTAGATATATTAACCTTCAATTTT 60
    |||
DB 1 GATCCATAACCCCAACCTATCGAAATATGAAATCTAGATATATTAACCTTCAATTTT 60
    |||

QY 61 TTAATATGTCGGTACGGGTACTGCTCTACCAAAACAGAGATTAACGACCGAC 120
    |||
DB 61 TTAATATGTCGGTACGGGTACTGCTCTACCAAAACAGAGATTAACGACCGAC 120
    |||

QY 121 CCTTTTAAGTCTTTTGAAGCTTACACCTGATTAACCGCTATGACGCCAGGACCGAA 180
    |||
DB 121 CCTTTTAAGTCTTTTGAAGCTTACACCTGATTAACCGCTATGACGCCAGGACCGAA 180
    |||

QY 181 TATTCGCGCCCTTACGCTCTGCTACGTCGATGATGATTCGTAAGTACCCGCTGCGC 240
    |||
DB 181 TATTCGCGCCCTTACGCTCTGCTACGTCGATGATGATTCGTAAGTACCCGCTGCGC 240
    |||

QY 241 AGGATAGTCTGGGATTTTGTATGATGATGATGATGATGATGATGATGATGATGAT 300
    |||
DB 241 AGGATAGTCTGGGATTTTGTATGATGATGATGATGATGATGATGATGATGATGAT 300
    |||

QY 301 GGCAGATGATGACCCGCGAGCGGTACACCTTTTGAAGTGGCGGGCGGCATATATTTGT 360
    |||
DB 301 GGCAGATGATGACCCGCGAGCGGTACACCTTTTGAAGTGGCGGGCGGCATATATTTGT 360
    |||

QY 361 CTGCGTGGCGCTGGAATAATCGCCACAGCCCAACAGGAAGACGACTTCGCGAAAC 420
    |||
DB 361 CTGCGTGGCGCTGGAATAATCGCCACAGCCCAACAGGAAGACGACTTCGCGAAAC 420
    |||

QY 421 AATCAGCTTTTGGGCGAGCTTGTGCTTGCATGTTGTGAAGCTCAAAATCATTTTGTACGG 480
    |||
DB 421 AATCAGCTTTTGGGCGAGCTTGTGCTTGCATGTTGTGAAGCTCAAAATCATTTTGTACGG 480
    |||

QY 481 TGTATCGGCACTGTCGACGTTTGTGCTGCGCAAAACAGAGCGTTAAAGCTGGTATGTTGG 540
    |||
DB 481 TGTATCGGCACTGTCGACGTTTGTGCTGCGCAAAACAGAGCGTTAAAGCTGGTATGTTGG 540
    |||

QY 541 CGTACGCGTTTGTGCTGCGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
    |||
DB 541 CGTACGCGTTTGTGCTGCGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
    |||

QY 601 GCATCTGTTTCAGGATGTTTTCGCGAGTATGTCGACGTTAAATATCGTTCGCTTGCCT 660
    |||
DB 601 GCATCTGTTTCAGGATGTTTTCGCGAGTATGTCGACGTTAAATATCGTTCGCTTGCCT 660
    |||

QY 661 GTTCTGCTATTTGCGCGGTACGCAATTTTCTATTAACGAAAAAAGCGAAGAGTGC 720
    |||
DB 661 GTTCTGCTATTTGCGCGGTACGCAATTTTCTATTAACGAAAAAAGCGAAGAGTGC 720
    |||

QY 721 CCTCTCGGCTTAACTTGTCTACTTAAG 750
    |||
DB 721 CCTCTCGGCTTAACTTGTCTACTTAAG 750
    |||

```

RESULT 2
 ID AAAS2690 standard; DNA; 588 BP.

XX AAAS2690;
 XX
 XX 03-JAN-2001 (first entry)
 XX
 DE Escherichia coli yfik gene.
 XX

KW E. coli; yfik gene; amino acid production; excretion protein gene;
 XX amino acid excretion protein; ds.
 XX

OS Escherichia coli.

EH Key Location/Qualifiers

FT CDS 1..588

FT /*tag= a

FT /product= "yfik"

XX EPI016710-A2.

XX 05-JUL-2000.

XX 17-DEC-1999; 99BP-00125263.

XX 30-DEC-1998; 98RU-00124016.

XX 09-MAR-1999; 99RU-00104431.

XX (AJIN) AJINOMOTO CO INC.

PI Livshits VA, Zakaraeva NP, Nakanishi K, Aleshin VV, Troshin PV;

PI Tokhmakova IV;

DR WPI; 2000-414802/36.

DR P-PSDB; AAB01788.

PT Increased production of L-amino acids by an Escherichia bacterium

PT comprises increasing the expression amount of an L-amino acid excretion

PT protein.

PS Disclosure; Page 22; 29pp; English.

XX The present sequence is the yfik gene (an excretion protein gene) of

CC Escherichia coli. The amino acid excretion protein produced from this

CC gene is involved in the production of amino acids, and an increase in its

CC expression leads to an increased accumulation of amino acids in the cell.

CC In this case, an increase in threonine, histidine, glutamic acid and

CC proline is achieved if multiple copies of the gene are transfected into a

CC bacterium. The bacterium used is E. coli

XX

SQ Sequence 588 BP; 113 A; 134 C; 167 G; 174 T; 0 U; 0 Other;

Query Match 78.4%; Score 588; DB 3; Length 588;

Best Local Similarity 100.0%; Pred. No. 1.1e-191;

Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 110 GTGACACCGAACCTTTTAAAGTCTTTTGAAGTTCACCCCTGATTAACCGCTATGACGCCA 169
    |||
DB 1 GTGACACCGAACCTTTTAAAGTCTTTTGAAGTTCACCCCTGATTAACCGCTATGACGCCA 169
    |||

QY 170 GGAACGAAACAATATTCGCGCCCTTAAGCTCTGCTACGTCGATGATGATTCGTAACGTAAC 229
    |||
DB 170 GGAACGAAACAATATTCGCGCCCTTAAGCTCTGCTACGTCGATGATGATTCGTAACGTAAC 229
    |||

QY 230 CGCGTCTGCGAGGATGAGTCTGCGATTTTGAATTTGATGATGATGATGATGATGATGAT 289
    |||
DB 230 CGCGTCTGCGAGGATGAGTCTGCGATTTTGAATTTGATGATGATGATGATGATGATGAT 289
    |||

QY 290 TCATTTTCACTGGCAGTATGACCCGCGACGCGTACACCTTTTGAATTTGATGATGATGAT 349
    |||
DB 290 TCATTTTCACTGGCAGTATGACCCGCGACGCGTACACCTTTTGAATTTGATGATGATGAT 349
    |||

QY 349 181 TCATTTTCACTGGCAGTATGACCCGCGACGCGTACACCTTTTGAATTTGATGATGATGAT 409
    |||
DB 349 181 TCATTTTCACTGGCAGTATGACCCGCGACGCGTACACCTTTTGAATTTGATGATGATGAT 409
    |||

QY 409 241 GCATATATTTGTCTGCGTGGCGTGAATAATGCCACCGCCCAACAAAGAAAGACGACTT 469
    |||
DB 409 241 GCATATATTTGTCTGCGTGGCGTGAATAATGCCACCGCCCAACAAAGAAAGACGACTT 469
    |||

QY 469 301 CAGGCAAAACCAATCAGCTTTTGGGCGAGCTTTGAGTTTGTGAAGTGAAGTCAAAATC 529
    |||
DB 469 301 CAGGCAAAACCAATCAGCTTTTGGGCGAGCTTTGAGTTTGTGAAGTGAAGTCAAAATC 529
    |||

```

Db 361 ATTGTGACGGGTGTACGGCAGTGTGACGTTGTCTCCGCAACACAGCGCTTAAGC 420
Qy 530 TGGGTAGTGGCGCTCAGCGTTTGTCTGCGCATGATGGAGCGTTGGCAATGTGTGCTGG 589
Db 421 TGGGTAGTGGCGCTCAGCGTTTGTCTGCGCATGATGGAGCGTTGGCAATGTGTGCTGG 480
Qy 590 GCGGTGCGGGGCGCATCTGTTTCAGCGATTTGTTGCGCAGTATGGTCCGCAATTAAATATC 649
Db 481 GCGGTGCGGGGCGCATCTGTTTCAGCGATTTGTTGCGCAGTATGGTCCGCAATTAAATATC 540
Qy 650 GTGCTTGCCTCTGTGCTGTGCTATTTGCGCGGTACGCAATTTCTATTAA 697
Db 541 GTGCTTGCCTCTGTGCTGTGCTATTTGCGCGGTACGCAATTTCTATTAA 588

RESULT 3
ID AAS81979 standard; cDNA; 1668 BP.
XX
AC AAS81979;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #17783.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR P-PSDB; ABG17792.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 17783; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping.
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polynucleotide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS8197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1668 BP; 376 A; 420 C; 445 G; 427 T; 0 U; 0 Other;
SQ
Query Match 39.4%; Score 295.8; DB 5; Length 1668;
Best Local Similarity 99.3%; Pred. No. 1.5e-90;
Matches 297; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 382 CACAGCCCAACAAAGGAACAGGACTTCAGGCAAAACAAATCAGCTTTGGCCAGCTT 441
Db 1014 CACTGCCCAACAAAGGAACAGGACTTCAGGCAAAACAAATCAGCTTTGGCCAGCTT 1073
Qy 442 TGCTTTCAGTTGTGTAACGTCAAAATCATTTTGTACGTGTACGCACTGTGACGTT 501
Db 1074 TGCTTTCAGTTGTGTAACGTCAAAATCATTTTGTACGTGTACGCACTGTGACGTT 1133
Qy 502 TGTTCCTCCGCAACACAGGCGTTTACCTGGGTAGTTGGCGTCAAGCTTTTGTCTGGCGAT 561
Db 1134 TGTTCTGCCCAACACAGGCGTTTACCTGGGTAGTTGGCGTCAAGCTTTTGTCTGGCGAT 1193
Qy 562 GATTGGGACGTTTGGCAATGTGTGCGGCGCTGGCGGCGGCGATCTGTTTCAGGCAATGTT 621
Db 1194 GATTGGGACGTTTGGCAATGTGTGCGGCGCTGGCGGCGGCGATCTGTTTCAGGCAATGTT 1253
Qy 622 TCGCCAGTATGTCGCGCAATTAAATATCGTCTTGGCCCTGTGCTGCTATTGGCGCG 680
Db 1254 TCGCCAGTATGTCGCGCAATTAAATATCGTCTTGGCCCTGTGCTGCTATTGGCGCG 1312

RESULT 4
ID ACH9695 standard; DNA; 618 BP.
XX
AC ACH9695;
XX
DT 29-JUL-2004 (first entry)
XX
DE Klebsiella pneumoniae polynucleotide seqid 5490.
XX
KM Recombinant expression vector; transcription regulatory element;
KM Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.
XX
OS Klebsiella pneumoniae.
XX
PN US6610836-B1.
XX
PD 26-AUG-2003.
XX
PF 27-JAN-2000; 2000US-00489039.
XX
PR 29-JAN-1999; 99US-0117747P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton GL, Osborne M;
XX
DR WPI: 2003-895346/82.
DR P-PSDB; ABO6144.
XX
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT preparing a vaccine composition against Klebsiella pneumoniae.
XX
PS Disclosure; SEQ ID NO 5490; 932pp; English.
XX
XX The invention describes a new isolated nucleic acid encoding a Klebsiella
XX pneumoniae polypeptide. Also described are: a recombinant expression
XX vector comprising the nucleic acid, operably linked to a transcription
XX regulatory element; and a cell comprising the recombinant expression
XX vector. The nucleic acid is useful for preparing a vaccine composition
XX against Klebsiella pneumoniae. This sequence encodes a Klebsiella
XX pneumoniae polypeptide of the invention
SQ
Sequence 618 BP; 110 A; 161 C; 183 G; 164 T; 0 U; 0 Other;

Query Match 35.4%; Score 265.8; DB 11; Length 618;
Best Local Similarity 65.0%; Pred. No. 1.9e-80;
Matches 393; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 83 TGTCTACCAAAACAGAGAGATTAACAAGTACACCGACCCCTTTTAAGTGTCTTTTGAAGT 142
1 TGTCTGCACAGCAAGAGAGAAACCGGTGACCTTACCTGATTAACGCGCTTTTAACT 60
DB 143 TACACCTGATTAACCGCTATGACCCGACGACCGAAACAATATTCGCGCTTACGTCT 202
61 TACACTTGTGATCACCGCCCTGACACCCGGGCCCAACATATCTGGCGCTGAGAGCGTC 120
QY 203 ACCGCGCATGATTTCCGCAAAAGTACCAGCGCTGCGAGGAGATGATCGAGATTTTGG 262
121 ACCCTCCATGATGTACACCGCGACGCTCGGGTGTGCGCGGATGACGTCGCGTTAT 180
QY 263 ATTGTGATGTTACTGTGTGCGGCAATTTCACTTTCAGTGCAGTATTAACCGCGACGC 322
181 ATCACTATGTTGATTTGCGCGGCTTTAACCTTTTCGCTGTTGAACCTGACTCAGCCTT 240
QY 323 GTACACCTTTGATTTGGCGGGCGGCATATATTTGCTGCGCTGCGGCAAAATGCGC 382
241 ACCCTGATTTGGATGATGATTTGGCGGCGCTGACATCTGCTGCGCTGCGCAATTTGC 300
QY 383 ACCAGCCCAACAAAGAAAGACGCACTTCAAGCAAAACCAATCAGCTTTTGGCGCAGCTT 442
301 AAAAGCAACCGGCGACAGGAGACCGCAATGTATTAACCGGTAGCTTTTGGCGAGTCTG 360
DB 443 GCTTTGACGTTTGTGAACGTCAAAATCATTTTGTACGCTGTATACGGCACTGTGACGTT 502
361 GGGCTACAGTTTGTGCAACGTTAAATATTTCTGATGACATCACCGGCTTTTCTAACCTT 420
QY 503 GTTGTGCGCGCAACACAGCGTTAAGCTGTGATTTGGCGTCAACCGTTTGTGCGCATG 562
421 GTTGTGCGCGCTTACTGCGAGCGCTTTTGTGCTTATACGAGTGTGCTGTGCGCGCG 480
DB 563 ATTGGACGTTTGGCAATGTGTGCTGCGCGCTGCGGCGCATCTGTTTACAGCATTTT 622
481 ATTGGCGCGCTGGAAATCTGTGCTGCGCGCTGCGGCGCATCTGTTTCAACGCGCTT 540
QY 623 CGCCAGTATGCTGCGCAGTTAATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 682
541 TTGCTGTACGGGCGGCAACTGAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
DB 683 CGCAT 687
601 CGCAT 605

RESULT 5
ACH9702
ID ACH9702 standard; DNA; 372 BP.
XX ACH9702;
AC
XX 29-JUL-2004 (first entry)
XX
XX Klebsiella pneumoniae polynucleotide seqid 5497.
DB
XX
XX Recombinant expression vector; transcription regulatory element;
KM Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.
XX
OS Klebsiella pneumoniae.
XX
XX US6610836-B1.
XX
XX 26-AUG-2003.
XX
XX 27-JAN-2000; 2000US-00489039.
XX
XX 29-JAN-1999; 99US-0117747P.
XX

PA (GENO-) GENOME THERAPEUTICS CORP.
PI Breton GL, Osborne M;
XX
XX WPI, 2003-895346/82.
DR P-PSDB; ABO66151.
XX
PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT preparing a vaccine composition against Klebsiella pneumoniae.
XX
PS Disclosure; SEQ ID NO 5497; 932pp; English.
XX
XX The invention describes a new isolated nucleic acid encoding a Klebsiella
CC pneumoniae polypeptide. Also described are: a recombinant expression
CC vector comprising the nucleic acid, operably linked to a transcription
CC regulatory element; and a cell comprising the recombinant expression
CC vector. The nucleic acid is useful for preparing a vaccine composition
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella
CC pneumoniae polypeptide of the invention
XX

SEQ Sequence 372 BP; 87 A; 95 C; 87 G; 103 T; 0 U; 0 Other;

Query Match 20.8%; Score 156; DB 11; Length 372;
Best Local Similarity 71.8%; Pred. No. 1.1e-42;
Matches 204; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 19 TATCGAAATATGAAATCTGAATATTAATAACATTCATTTTAAATGTTCCGTGCG 78
85 TATGATATTTATGAAAGCGAATGAAATATATCTATTTTAAATATTCAGCTTCGG 144
DB 79 GTAAGTGTCTACCAAAACAGAGAGATTAACAAGTACACCGACCCCTTTTAAGTCTTTTGG 138
145 GTAAGTGTCTACCAAGAGAGAAACCGGTGACACCTGATTAAGCGCTTTT 204
QY 139 GACTTACACCTGATTAACGCTATGACCGCAGACCGAATAATTTCTGCCCTTATGCTC 198
205 AACTTACACTTATATCACCGCCCTGACACCGCGGCCCAATAATCTGTGCGCTGAGCAG 264
DB 199 TGCTAGTGTGATGATTTGTCAAAAGTACCCGCGTGTGCGAGGATGATGCTGCGATT 258
265 GGTACCTCCCAAGGTGTACGCGCAGACCTGCGGAGTCTGCGGAGTATGCGGT 324
QY 259 TTTGATTTGATGATTTACTGTGTGCGGCAATTTTCACTGG 302
325 TATTATCATATGATGATTTGCGCGGCTTTAACCTTTTTCCTG 368
DB

RESULT 6
ACD79726
ID ACD79726 standard; DNA; 100 BP.
XX
XX ACD79726;
AC
XX 19-SEP-2003 (first entry)
XX
XX E. coli K12 MG1655 biochip probe SEQ ID 11002.
DB
XX
XX Biochip; gene expression; gut; diagnostic; detection; probe; ss.
KM
XX
XX Escherichia coli.
OS
XX
XX EPI260592-A1.
XX
XX 27-NOV-2002.
XX
XX 17-MAY-2001; 2001EP-00112179.
XX
XX 17-MAY-2001; 2001EP-00112179.
XX
XX (MMGB-) MMG-BIOTECH AG.
XX
XX Donner H, Drescher B, Huber A, Weber J;
XX

DR WPI; 2003-241155/24.
XX Blochip containing probes complementary with open reading frames in
PT Escherichia coli K12, useful for detecting gene expression and expression
PS patterns.
PS Claim 3; Page 1714; 2004pp; German.
XX
CC This invention describes a novel biochip comprising probe spots, each
CC containing many identical probes. The probes are nucleotide sequences of
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC least one includes a segment of at least 20 bases identical with, or
CC complementary to, a segment of an open reading frame (orf) of Escherichia
CC coli K12. The biochip is used for specific detection of gene expression
CC in K12 and for determining the gene expression pattern, e.g. for
CC diagnostic determination of which E. coli strains are present in the gut,
CC and to determine the effects of e.g. growth media on gene expression. The
CC biochip provides as comprehensive as possible detection of the K12
CC genome, with simultaneous analysis of many different genes with a single
CC device, and comparison of gene expression between K12 and its mutants or
CC other E. coli strains in a single experiment. Apart from qualitative and
CC quantitative information about gene expression, it also allows
CC measurements of population densities for the various strains. The use of
CC synthetic oligonucleotides for preparation of probes allows free
CC variation in probe length and ensures high purity (and thus selectivity,
CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC ACD81540 represent oligonucleotide probes used with the biochip described
CC in the invention
XX
SQ Sequence 100 BP; 18 A; 20 C; 30 G; 32 T; 0 U; 0 Other;
XX
Query Match 13.3%; Score 100; DB 8; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.1e-23;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 466 AATCATTTTGTACGGTGTACGACGACGCTTGTCTGCGCAACACAGGCGTT 525
DB 1 AATCATTTTGTACGGTGTACGACGACGCTTGTCTGCGCAACACAGGCGTT 60
QY 526 AAGCTGGGTAGTGGCGCTCAAGCGTTTGTCTGCGCATGATT 565
DB 61 AAGCTGGGTAGTGGCGCTCAAGCGTTTGTCTGCGCATGATT 100
RESULT 7
ACD79725
ID ACD79725 standard; DNA; 100 BP.
XX
AC ACD79725;
XX
DT 19-SEP-2003 (first entry)
XX
DE E. coli K12 MG1655 biochip probe SEQ ID 11001.
XX
KM Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS Escherichia coli.
XX
PN EP1260592-A1.
XX
PD 27-NOV-2002.
XX
PF 17-MAY-2001; 2001EP-00112179.
XX
PR 17-MAY-2001; 2001EP-00112179.
XX
PA (MMGB-) MMGB-BIOTECH AG.
XX
PI Donner H, Dreischer B, Huber A, Weber J;
XX
DR WPI; 2003-241155/24.
XX

PT Blochip containing probes complementary with open reading frames in
PT Escherichia coli K12, useful for detecting gene expression and expression
PS patterns.
PS Claim 3; Page 1714; 2004pp; German.
XX
CC This invention describes a novel biochip comprising probe spots, each
CC containing many identical probes. The probes are nucleotide sequences of
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC least one includes a segment of at least 20 bases identical with, or
CC complementary to, a segment of an open reading frame (orf) of Escherichia
CC coli K12. The biochip is used for specific detection of gene expression
CC in K12 and for determining the gene expression pattern, e.g. for
CC diagnostic determination of which E. coli strains are present in the gut,
CC and to determine the effects of e.g. growth media on gene expression. The
CC biochip provides as comprehensive as possible detection of the K12
CC genome, with simultaneous analysis of many different genes with a single
CC device, and comparison of gene expression between K12 and its mutants or
CC other E. coli strains in a single experiment. Apart from qualitative and
CC quantitative information about gene expression, it also allows
CC measurements of population densities for the various strains. The use of
CC synthetic oligonucleotides for preparation of probes allows free
CC variation in probe length and ensures high purity (and thus selectivity,
CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC ACD81540 represent oligonucleotide probes used with the biochip described
CC in the invention
XX
SQ Sequence 100 BP; 16 A; 22 C; 26 G; 36 T; 0 U; 0 Other;
XX
Query Match 13.3%; Score 100; DB 8; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.1e-23;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 598 GGGGCATCTGTTTACGAGATTTGTTGCGCAGTATGTCGCGCAGTTAAATTCGTGTC 657
DB 1 GGGGCATCTGTTTACGAGATTTGTTGCGCAGTATGTCGCGCAGTTAAATTCGTGTC 60
QY 658 CCGTTCCTGCTTATTTGCGCGGTACGCAATTTTCTATTAA 697
DB 61 CCGTTCCTGCTTATTTGCGCGGTACGCAATTTTCTATTAA 100
RESULT 8
ACD79727
ID ACD79727 standard; DNA; 100 BP.
XX
AC ACD79727;
XX
DT 19-SEP-2003 (first entry)
XX
DE E. coli K12 MG1655 biochip probe SEQ ID 11003.
XX
KM Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS Escherichia coli.
XX
PN EP1260592-A1.
XX
PD 27-NOV-2002.
XX
PF 17-MAY-2001; 2001EP-00112179.
XX
PR 17-MAY-2001; 2001EP-00112179.
XX
PA (MMGB-) MMGB-BIOTECH AG.
XX
PI Donner H, Dreischer B, Huber A, Weber J;
XX
DR WPI; 2003-241155/24.
XX
PT Blochip containing probes complementary with open reading frames in
PT Escherichia coli K12, useful for detecting gene expression and expression

PT patterns.
XX
PS Claim 3; Page 1714; 2004pp; German.
XX
CC This invention describes a novel biochip comprising probe spots, each
CC containing many identical probes. The probes are nucleotide sequences of
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC least one includes a segment of at least 20 bases identical with, or
CC complementary to, a segment of an open reading frame (orf) of *Escherichia*
CC coli K12. The biochip is used for specific detection of gene expression
CC in K12 and for determining the gene expression pattern, e.g. for
CC diagnostic determination of which *E. coli* strains are present in the gut,
CC and to determine the effects of e.g. growth media on gene expression. The
CC biochip provides a comprehensive as possible detection of the K12
CC genome, with simultaneous analysis of many different genes with a single
CC device, and comparison of gene expression between K12 and its mutants or
CC other *E. coli* strains in a single experiment. Apart from qualitative and
CC quantitative information about gene expression, it also allows
CC measurements of population densities for the various strains. The use of
CC synthetic oligonucleotides for preparation of probes allows free
CC variation in probe length and ensures high purity (and thus selectivity,
CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACD8731 to
CC ACD81540 represent oligonucleotide probes used with the biochip described
CC in the invention
XX
SQ Sequence 100 BP; 23 A; 22 C; 25 G; 30 T; 0 U; 0 Other;
Query Match 13.3%; Score 100; DB 8; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.1e-23;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DY 401 GACGAGCTTCAGGAAACCAATCAGCTTTGGGCGAGCTTTGCTTGACGTTGTGAC 460
DB 1 GACGAGCTTCAGGAAACCAATCAGCTTTGGGCGAGCTTTGCTTGACGTTGTGAC 60
DY 461 GTCAAAATCATTTGTGTACGCGTGTACGCACTGTGACGT 500
DB 61 GTCAAAATCATTTGTGTACGCGTGTACGCACTGTGACGT 100
RESULT 9
ABK73143
ID ABK73143 standard; DNA; 573 BP.
XX
AC ABK73143;
XX
DT 13-AUG-2002 (first entry)
XX
DE *Bacillus licheniformis* genomic sequence tag (GST) #434.
XX
KW Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.
XX
OS *Bacillus licheniformis*.
XX
PN WO200229113-A2.
XX
PD 11-APR-2002.
XX
PF 05-OCT-2001; 2001WO-US031437.
XX
PR 06-OCT-2000; 2000US-00680598.
PR 27-MAR-2001; 2001US-0279526P.
XX
PA (NOVO) NOVOZYMES BIOTECH INC.
PA (NOVO) NOVOZYMES AS.
XX
PI Berka R, Clausen IG;
XX
DR WPI; 2002-416684/44.
XX

PT Monitoring differential expression of several genes in first *Bacillus*
PT cell relative to expression of same genes in one or more second *Bacillus*
PT cells, by using substrate containing *Bacillus* genomic sequenced tag
PT array.
XX
PS Claim 4; SEQ ID NO 434; 200pp; English.
XX
XX
CC The invention describes a method of monitoring differential expression of
CC genes in a first *Bacillus* cell relative to expression of the genes in
CC other *Bacillus* cells, comprising hybridising labelled nucleic acid probes
CC isolated from *Bacillus* cells to a substrate containing array of *Bacillus*
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first *Bacillus* cell relative to expression of the same genes
CC in one or more second *Bacillus* cells. The method is useful for monitoring
CC global expression of several genes from a *Bacillus* cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which *Bacillus* cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive follow
CC -up characterisation is unnecessary, when one spot on an array equals one
CC gene or one open reading frame, since sequence information is available.
CC This sequence represents a genomic sequence tag (GST) used in the method
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 573 BP; 128 A; 116 C; 129 G; 200 T; 0 U; 0 Other;
Query Match 6.7%; Score 50.4; DB 6; Length 573;
Best Local Similarity 46.6%; Pred. No. 4.2e-06;
Matches 194; Conservative 0; Mismatches 221; Indels 1; Gaps 1;
DY 124 TTAAAGTCTTTTGAAGCTTACACCTGATTACCGCTATGACCGCAAGCAATAT 183
DB 6 TATTATAGCTTTTCTTTCGTATGTGATTATGACATCATTAACCCGCGCGCAATAT 65
DY 184 TCTCGCCCTTAGCTCTGATACGTCGATGATTTGCTCAAGTACCGCGTGTGACG 243
DB 66 TTTAATGATGATGAAGAGCGAAAGTTGCGCTTTACAGGTTACGCGTTTACAGCG 125
DY 244 GATGAGCTGCGAATTTTGTGATGTGATGATGATGATGATGATGATGATGATGATG 303
DB 126 TATCTTGGCGGGGTTTGCATCTTGGGATCTCAGCGGTGCTTTACATAGCGCTGA 185
DY 304 AGTGAATGACCCGCGAGCGGTACACCTTTTGAAGTTGGCGGGCGGCGATATATG 363
DB 186 CAATYGATTCGCCGTGTAGAGCTTATTTTAAACTTGGCGGCGCGGTATATGATTTA 245
DY 364 GCTGCGCTGGAATTAATGCCACAGCCCAAGAAAGAAAGAGAGCTTACGCAAAACAAT 423
DB 246 TTTGGCTTTGCAAGTGGCTTTTACCAAAATTAAGAGATTCACAGAACCGCGCTC 305
DY 424 CAGCTTTTGGCGAGCTTTGCTTTGCAAGTTTGAAGCTTCAATATTTTGTACGCTGT 483
DB 306 CTCTTTTATATCCGGCTTTATTTTCAAGCTGATCAATATTAAGAGATTTTGTCTCAT 365
DY 484 T-ACGCACTGTGACGTTTGTCTGCGCAACACAGCGCTTAAAGCTGGGTGTT 538
DB 366 TAACCGTATGAGCGCGTGTGCTTGGCGCTTCAACATTCCTGAATTCAGCAATT 421
RESULT 10
AAS92926
ID AAS92926 standard; cDNA; 2325 BP.
XX
AC AAS92926;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #28730.
XX

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS WO200175067-A2.
PN WO200175067-A2.
XX 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US008631.
PF 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
PR (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX P-PSDB; ABG28735.
DR New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 1; SEQ ID NO 28730; 103bp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 2325 BP; 510 A; 570 C; 642 G; 603 T; 0 U; 0 Other;
SQ
Query Match 5.7%; Score 42.6; DB 5; Length 2325;
Best Local Similarity 53.7%; Pred. No. 0.0046;
Matches 110; Conservative 0; Mismatches 94; Indels 1; Gaps 1;
QY 465 AATCATTTTGTACGGTGTACGCACTGTGACGTTTGTCCGCCAACAACAGCGT 524
DB 440 AACTTTTTCCTTGCGGGGTAGGACATCCGTGCGCAATTTTCTCAATAATGAGACT 499
QY 525 TAAGCTGGGTAGTGGCGGTACGCTTTTCTGCGGATGATGGAGACGTTTGGCAATGTG 584
DB 500 GAAAACTCTGCTCAAGCGTTTTCCTTGCGGATGATGGAGCTTTTGGAAAAATGT 559
QY 585 GCTGGGCGCTGGCGGGGATCTGTTTCAAGCATTTGTTTCCGCAATGATGCGCAGTTA 643
DB 560 CTGGGGGCTGGCGGGGATTTTTCATGAATTTTTCCTCCATTAAGTCCCATTA 619
QY 644 AATATCGTCTGCTGCTGTTGCTGG 668
DB 620 AATATCGGGGGTTCCTGCTGCTGG 644

RESULT 11
AAS88917/C
ID AAS88917 standard; cDNA; 3020 BP.
XX
AC AAS88917;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #24721.
XX
KM Human: chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
PN WO200175067-A2.
XX
PD 11-OCT-2001.
PF 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX P-PSDB; ABG24730.
DR New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 1; SEQ ID NO 24721; 103bp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 3020 BP; 852 A; 763 C; 758 G; 647 T; 0 U; 0 Other;
SQ
Query Match 5.7%; Score 42.6; DB 5; Length 3020;
Best Local Similarity 53.7%; Pred. No. 0.0054;
Matches 110; Conservative 0; Mismatches 94; Indels 1; Gaps 1;
QY 465 AATCATTTTGTACGGTGTACGCACTGTGACGTTTGTCCGCCAACAACAGCGT 524
DB 1888 AACTTTTTCCTTGCGGGGTAGGACATCCGTGCGCAATTTTCTCAATAATGAGACT 1829
QY 525 TAAGCTGGGTAGTGGCGGTACGCTTTTCTGCGGATGATGGAGACGTTTGGCAATGTG 584

```
Db      1828 GAAACTCTGCTCAAGCGTTTTTCCCTGGCGATATTGNGAGACTTTGGAAAAATGTT 1769
Qy      585 GCTGGGCGCTGGCGGGGCAATCTGTTTCAGCGATTGTTTC-GCCAGTATGCTCCGCACTTA 643
      1768 CTTGGGGGCTGGCGGGGGAATATTTTTCACGAATTTTCTCCAAATAGGTCCCAATTA 1709
Qy      644 AATATCGTGCCTGGCCCTGTTGCTGG 668
      1708 AATCGGGGCTTCCCTGTTGCTGG 1684

RESULT 12
ADA31058
ID      ADA31058 standard; DNA; 669 BP.
AC      ADA31058;
XX
XX      20-NOV-2003 (first entry)
XX
DE      DNA encoding Acinetobacter baumannii protein #2345.
XX
XX      ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
KM      vaccine; plant biocontrol agent.
XX
OS      Acinetobacter baumannii.
XX
XX      US6562958-B1.
XX
XX      13-MAY-2003.
XX
XX      04-JUN-1999; 99US-00328352.
XX
XX      09-JUN-1998; 98US-0088701P.
XX
PA      (GENO-) GENOME THERAPEUTICS CORP.
PI      Breton G, Bush D;
PI
PI      WPI; 2003-576092/54.
DR      P-PSDB; ADA35184.
XX
XX      New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT      for diagnosing a bacterial disease, as components of antibacterial
PT      vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT      plants.
XX
XX      Example; SEQ ID NO 2345; 328bp; English.
XX
PS      The invention relates to isolated Acinetobacter baumannii nucleic acids.
XX
XX      The A. baumannii nucleic acids and polypeptides are useful as reagents
CC      for diagnosing a bacterial disease, as components of antibacterial
CC      vaccines, as targets for antibacterial drugs, to detect the presence of
CC      A. baumannii and other Acinetobacter species in a sample, in screening
CC      compounds for the ability to interfere with the A. baumannii life cycle
CC      or to inhibit A. baumannii infection, and as biocontrol agents for
CC      plants. The present sequence represents DNA encoding an A. baumannii
CC      protein.
XX
XX      Sequence 669 BP; 152 A; 133 C; 155 G; 229 T; 0 U; 0 Other;
SQ
Query Match      5.6%; Score 42; DB 9; Length 669;
Best Local Similarity 50.4%; Pred. No. 0.0036;
Matches 133; Conservative 0; Mismatches 125; Indels 6; Gaps 1;
```

```
Db      167 TCGTCCCCGCTGTGAGGATGCTATTCGGGCACTTATAGATTTATTGTAAGCGG 226
Qy      285 GCATTTCAATTTTACATCGGACGATGATGACCCGGCAGCGGTACACCTTTGAGTGGCCG 344
      227 CTGGTGTGGTCTGCTGTGATGCAAAATPACCACTGCTGTGATGATTTACTGTTGCCG 286
Qy      345 GGGCGGCATATATATGCTGAGCTGG 368
      287 GTTCTGCTATTATATATGATGCG 310

RESULT 13
ADF02251
ID      ADF02251 standard; DNA; 636 BP.
AC      ADF02251;
XX
XX      12-FEB-2004 (first entry)
XX
DE      Bacterial polynucleotide #2536.
XX
XX      Proteus mirabilis infection; bacterial infection; antibacterial;
KM      immunostimulant; gene; ds.
XX
XX      Proteus mirabilis.
XX
XX      US6605709-B1.
XX
XX      12-AUG-2003.
XX
XX      05-APR-2000; 2000US-00543681.
XX
XX      09-APR-1999; 99US-0128706P.
XX
PA      (GENO-) GENOME THERAPEUTICS CORP.
PI      Breton GL;
PI
PI      WPI; 2003-895291/82.
DR      P-PSDB; ADF06423.
XX
XX      New Proteus mirabilis polypeptides and polynucleotides, useful as
PT      reagents for diagnosis of bacterial disease, as components of
PT      antibacterial vaccines, as targets for antibacterial drugs, or as
PT      biocontrol agents for plants.
XX
XX      Disclosure; SEQ ID NO 2536; 870bp; English.
XX
PS      The invention relates to new Proteus mirabilis polypeptides and
XX
XX      polynucleotides. The invention also relates to antibodies against the
CC      polypeptides, methods for producing the polypeptides, a method of
CC      generating vaccines for immunising an individual against P. mirabilis, a
CC      method for evaluating a compound for the ability to bind a P. mirabilis
CC      polypeptide and a method for screening test compounds for anti-bacterial
CC      activity. The polypeptides and polynucleotides are useful as molecular
CC      targets for diagnosing, preventing and treating pathological conditions
CC      resulting from bacterial infection, as reagents for diagnosis of
CC      bacterial diseases, as components of antibacterial vaccines, as targets
CC      for antibacterial drugs or as bio-control agents for plants. This
CC      sequence represents a Proteus mirabilis polynucleotide of the invention.
XX
XX      Sequence 636 BP; 158 A; 97 C; 144 G; 237 T; 0 U; 0 Other;
SQ
Query Match      5.3%; Score 39.8; DB 10; Length 636;
Best Local Similarity 46.5%; Pred. No. 0.02;
Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
```

```
Qy      111 TGACACGACCCCTTTAAGCTTTTGGACTTACACCCGTATGACGCGCAG 170
      32 TGACAATATACCTTACTGTTTCTTGGCGCTTTCCCTTTTATGACGCAATACGCGCG 91
Db      171 GACCGAACAATATTTCTGCGCCTTAGCTCTGCTACGTCGATGATTTTCGTAAGTACC 230
Qy
```



```

Db      92  GACCAATTAATTACTATTACCTCTTGGGGGCTCAGTTGGTTAAAGCCTCTTAA 151
Qy      231 GCGTCTGGCAGGAGTGAAGTCTGGAGTTTGTGATTTGATGATCTGTCGGGCAATT 290
Db      152 CGTGAATGGCAGGAGTATTCCTCGAATGCAATGTGTATCTTAGCTCTCGTTGGTA 211
Qy      291 CATTTCACCTGGCAGTGAATGACCCGGCAGCGGTACACCTTTGAGTTGGGCGGGCGG 350
Db      212 TTGGCAGATTATTTGATTATTTATCTCGATTGCAATCGTTTAAAGATTATTTGTAACA 271
Qy      351 CATATATTGTCGTGGCTGGCGTGAATAATCGCCACC 385
Db      272 TTATCTTTGTTGGTTAGCTTGAAACAGCAACC 306

RESULT 14
ABD10542 ID ABD10542 standard; DNA; 750 BP.
XX
AC ABD10542;
XX
DT 29-JUL-2004 (first entry)
XX
DE Pseudomonas aeruginosa polymnucleotide #9146.
XX
KM Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
XX antibacterial.
XX
OS Pseudomonas aeruginosa.
XX
PN US6551795-B1.
XX
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252391.
XX
PR 18-FEB-1998; 98US-0074788P.
XX
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI: 2003-615309/58.
XX
DR P-PSDB; ABO76971.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
XX useful as molecular targets for diagnostics, prophylaxis and treatment of
XX pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 9146; 455bp; English.
XX
CC The invention relates to Pseudomonas aeruginosa polypeptides and the
XX polymnucleotides encoding them. The sequences are useful in diagnosis and
XX therapy of pathological conditions, as molecular targets for diagnostics,
XX prophylaxis and treatment of pathological conditions resulting from a
XX bacterial infection, for evaluating a compound, such as a polypeptide,
XX for the ability to bind a P. aeruginosa nucleic acid, as components of
XX effective antibacterial targets, as targets for antibacterial drugs,
XX including anti-P. aeruginosa drugs, as templates for recombinant
XX production of P. aeruginosa-derived peptides or polypeptides, as target
XX components for diagnosis and/or treatment of P. aeruginosa-caused
XX infection, and in detection of P. aeruginosa sequences or other sequences
XX of Pseudomonas species using biotech technology. Sequences ABD10542-
XX ABD10547 represent P. aeruginosa polymnucleotides of the invention. Note:
XX The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format from USPTO at
XX seqdata.uspto.gov/sequence.html
XX
SQ Sequence 750 BP; 113 A; 232 C; 248 G; 157 T; 0 U; 0 Other;
Query Match 5.3%; Score 39.6; DB 11; Length 750;

```

```

Best Local Similarity 47.2%; Pred. No. 0.026;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;
Qy      131 GCTTTTGGACTTACACCTGATTAACCGCTATGACGCGACAGCAACAAATATTCGCC 190
Db      160 GCGTTCCTCCGTTGCTTCGCTTCACCTCGTACGCGCCGGTCCGAACACCATGTGG 219
Qy      191 CTTAGCTCTGCTACGCTGATGATGATTTCCGAAAGTACCCGCTGCTGGCAGGATGAT 250
Db      220 CTGCGCTCGGGGGGTCAACTTCGGTTTCGTCCTCCATCCGACATCTCGGGATCAGT 279
Qy      251 CTGGGATTTTGTGATTTGATGATTTACTGTCGCGGCAATTCATTTCACTGCAATGAT 310
Db      280 TGGCGCTTCTTCACTGATGATGATGCGGTGGCTTCGCGCTGGGACGATGTCAGGCC 339
Qy      311 GACCCGGCAGCGGTACACCTTTGATTTGGCGGGCGGCGCATATATTTGTCGTGGCGCG 370
Db      340 TATCCGGTGTCTGTATACCATCTGCGCTACGTGCGGCGGCGCTACCTGCTGACCTGGCC 399
Qy      371 TGGAAATTCGCCAC 384
Db      400 TGGAAATTCGCCAC 413

RESULT 15
ADL08109 ID ADL08109 standard; DNA; 247682 BP.
XX
AC ADL08109;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human gene associated with low HDL-C A73.
XX
KM Human; ds; SNP; single nucleotide polymorphism;
XX high density lipoprotein-C; HDL-C; vascular disease; metabolic disease;
XX coronary artery disease; gene.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
XX FT Variation /*tag= a
XX FT /standard_name= "Single nucleotide polymorphism"
XX
PN US2004043389-A1.
XX
PD 04-MAR-2004.
XX
PR 04-SEP-2002; 2002US-00235192.
XX
PR 04-SEP-2002; 2002US-00235192.
XX
PA (VITI-) VITIVIFY INC.
XX
PI McCarthy J;
XX
DR WPI: 2004-214170/20.
XX
PT Determining whether a subject has, or is at risk of developing, an
XX abnormally low high density lipoprotein-C (HDL-C) level comprises
XX detecting an allelic variant of a polymorphic region from any of a set of
XX 27 genes.
XX
PS Disclosure; SEQ ID NO 28; 37bp; English.
XX
CC The invention relates to determining whether a subject has, or is at risk
XX of developing, an abnormally low high density lipoprotein-C (HDL-C) level
XX comprises determining whether the subject has an allelic variant of a
XX polymorphic region from any of 27 genes (alleles listed in Table 5 of the
XX specification). Also included are determining whether a male subject has,
XX or is at risk of developing, an abnormally low HDL-C level, comprising
XX determining whether the male subject has an allelic variant of a

```

CC polymorphic region listed in Table 5 which is associated with abnormally
CC low HDL-C levels in males, and determining whether a female subject has,
CC or is at risk of developing, an abnormally low HDL-C level, comprising
CC determining whether the female subject has an allelic variant of a
CC polymorphic region listed in Table 5 which is associated with abnormally
CC low HDL-C levels in females. The allelic variant in determining whether a
CC subject has, or is at risk of developing, an abnormally low HDL-C level
CC is APOA 1 CC, CD14 1 CT, COL5A2 1 GG, EDNRB 1 AG or AA, FABP3 1 CT, GBE1
CC 1 AG or GG, LIPC 5 AA, MTHFR 1 CC, VWF 2 GG, or their complements. The
CC allelic variant in determining whether a male subject has, or is at risk
CC of developing, an abnormally low HDL-C level, LRP1 3 CC or CT, PAI2 4 GG,
CC or PPARG 1 CC, or their complements. The allelic variants are also COL5A2
CC 1 GG, CD14 1 CT or CC, and FABP3 1 CT, in combination, or their
CC complements. The methods are useful for diagnosing (a predisposition to)
CC abnormally low levels of low high density lipoprotein-C (HDL-C) in a
CC subject, The methods are useful in diagnosing (a predisposition to) or
CC prognosticating diseases and disorders associated with abnormal lipid
CC levels such as vascular and metabolic diseases, e.g., coronary artery
CC disease. The present sequence is a human gene containing a SNP (single
CC nucleotide polymorphism associated with low high density lipoprotein-C
CC (HDL-C) levels.

XX Sequence 247682 BP; 75816 A; 50170 C; 48727 G; 72669 T; 0 U; 300 Other;

Query Match 5.2%; Score 39; DB 12; Length 247682;

Best Local Similarity 55.6%; Pred. No. 1.2;

Matches 75; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 6 ATAAACCCCAACCTATCGAAAATATCGAATCTAGAAATATAAAAACATTCAATTTTAA 65

Db 168128 ATAAACCTATATCTATGAAATTAATGATTAATTAATTAACCTTTCACAAAGAAA 168187

QY 66 TGTTCCTGTGTCGTACTGTGTACCAAAACGAGAGATTAACAAGTGACACCGACCTTT 125

Db 168188 ATGCCAGTCCCGAGTGGTGAATTTATCAAAAATGTAAAGAAATATAACATAATCTAT 168247

QY 126 TAAGTGCTTTTGA 140

Db 168248 GCAAAATCTTTCAG 168262

Search completed: September 10, 2005, 07:59:20
Job time : 479 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2005, 01:31:52 ; Search time 166 Seconds
(without alignments)
454.327 Million cell updates/sec

Title: US-10-620-487-2

Perfect score: 996
Sequence: 1 VTPTLSAFWTYTLITAMP.....RQNLIVALLLVCAVRIF 195

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	996	100.0	195	8	ADK66006
2	993	99.7	195	3	ABO1788
3	684	68.7	205	7	ABO66144
4	505	50.7	344	4	ABG17792
5	272.5	27.4	249	7	ABO76971
6	256	25.7	203	6	ABM70439
7	250.5	25.2	211	7	ABO63925
8	232.5	23.3	212	7	ADFO4482
9	228.5	22.9	211	7	ADFO6423
10	191.5	19.2	238	6	ADA35835
11	178	17.9	241	6	ADA35872
12	173	17.4	271	7	ABO83709
13	163.5	16.4	235	6	ADA36962
14	162	16.3	214	7	ABO61549
15	161.5	16.2	206	6	ABU32233
16	158.5	15.9	211	6	ABO70000
17	158.5	15.8	211	6	ABU39959
18	157.5	15.8	190	4	ABG29208
19	157.5	15.8	206	6	ABU28287
20	156.5	15.7	206	6	ABU50100
21	156.5	15.7	206	6	ABU47858
22	155.5	15.6	214	6	ADA34738
23	155.5	15.6	237	7	ABO78429
24	154	15.5	205	6	ABU44939
25	154	15.5	210	6	ADA35521

26	153.5	15.4	211	7	ABO72001	ABO72001	Pseudomon
27	153	15.4	214	7	ABO63276	ABO63276	Klebsiell
28	151.5	15.2	211	6	ABU9902	ABU9902	Protein e
29	149.5	15.0	250	7	ADFO4202	ADFO4202	Bacterial
30	148	14.9	211	7	ABO63091	ABO63091	Klebsiell
31	146.5	14.7	220	6	ADA3616	ADA3616	Acinetoba
32	146	14.7	222	6	ADA35184	ADA35184	Acinetoba
33	143.5	14.4	212	3	ABO1787	ABO1787	Escherich
34	138.5	13.9	219	7	ABO61214	ABO61214	Klebsiell
35	137.5	13.8	205	6	ABU99040	ABU99040	Protein e
36	132	13.3	208	6	ADA33824	ADA33824	Acinetoba
37	129	13.0	201	6	ABM68999	ABM68999	Phototrab
38	128.5	12.9	228	7	ADFO4569	ADFO4569	Bacterial
39	128.5	12.9	240	6	ADA36250	ADA36250	Acinetoba
40	128	12.9	220	7	ADFO4260	ADFO4260	Bacterial
41	127	12.8	200	6	ADA33486	ADA33486	Acinetoba
42	126	12.7	205	3	AAV79298	AAV79298	E. coli R
43	126	12.7	205	3	AAV9597	AAV9597	E. coli R
44	125.5	12.6	214	6	ADA35408	ADA35408	Acinetoba
45	123.5	12.4	209	6	ABU38890	ABU38890	Protein e

ALIGNMENTS

RESULT 1
ADK66006 standard; protein; 195 AA.
ID ADK66006

AC ADK66006;
XX
DT 06-MAY-2004 (first entry)
XX
DE E coli yfik protein.
XX
KM yfik; phosphoglycerate amino acid; amino acid production.
XX
OS Escherichia coli.
XX
FN EPI382684-A1.
XX
PD 21-JAN-2004.
XX
PF 10-JUL-2003; 2003EP-00015546.
XX
PR 19-JUL-2002; 2002DE-01032930.
XX
PA (CONE) CONSORTIUM ELEKTROCHEM IND GMBH.
XX
PI Mater T;
XX
DR WPI; 2004-192974/19.
XX
DR N-PSDB; ADK66005.
XX
PT New microbial strain, useful for producing amino acids of the
PT phosphoglycerate family, particularly serine and cysteine, has increased
PT activity of the yfik gene product.
XX
PS Disclosure; Page 10-11; 16pp; German.
XX
CC The present invention relates to a microbial strain for fermentative
CC product of amino acids of the phosphoglycerate family, or their
CC derivatives, has higher activity of the product of the yfik gene, or its
CC homologue, compared with its parent strain. The microbial strain is used
CC to produce O- and N-acetyl-serine, L-cysteine and cystine. The present
CC sequence is the E. coli yfik protein.
XX
SQ Sequence 195 AA;

Query Match 100.0%; Score 996; DB 8; Length 195;
Best Local Similarity 100.0%; Pred. No. 1,4e-109;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 VTPTLLSAFWTYTLITAMTPGPNNTILASATSHGFRQSTRVLAGMSLGLIYVLLCAGI 60
   |||||
Db 1 VTPTLLSAFWTYTLITAMTPGPNNTILASATSHGFRQSTRVLAGMSLGLIYVLLCAGI 60

QY 61 SFSILAVIDPAVAHLLSWAGAAYIYVWLAKIATSPTEKEDGIQAKPISFWASFPALQFVNVKI 120
   |||||
Db 61 SFSILAVIDPAVAHLLSWAGAAYIYVWLAKIATSPTEKEDGIQAKPISFWASFPALQFVNVKI 120

QY 121 ILVGTALSTFVLPTQOALSWMVGVSVLLAMIGTFGNVCWALAGHLFORLFRQYGRQNT 180
   |||||
Db 121 ILVGTALSTFVLPTQOALSWMVGVSVLLAMIGTFGNVCWALAGHLFORLFRQYGRQNT 180

QY 181 VIALLLVYCAVRIFY 195
   |||||
Db 181 VIALLLVYCAVRIFY 195

RESULT 2
AAB01788 standard; protein; 195 AA.
ID AAB01788
AC AAB01788;
XX
XX 03-JAN-2001 (first entry)
XX
XX Escherichia coli yfjK amino acid excretion protein.
XX
XX E. coli; yfjK gene; amino acid production; excretion protein gene;
XX
XX amino acid excretion protein.
XX
XX Escherichia coli.
XX
XX EPI016710-AA.
XX
XX 05-JUL-2000.
XX
XX 17-DEC-1999; 99EP-00125263.
XX
XX 30-DEC-1998; 98RU-00124016.
XX
XX 09-MAR-1999; 99RU-00104431.
XX
XX (AJIN ) AJINOMOTO CO INC.
XX
XX Iivshits VA, Zakaraeva NP, Nakanishi K, Aleshin VV, Troshin PV,
XX
XX Tokmakova IL;
XX
XX WPI; 2000-414802/36.
XX
XX N-PSDB; AAA52690.
XX
XX Increased production of L-amino acids by an Escherichia bacterium
XX
XX PT comprises increasing the expression amount of an L-amino acid excretion
XX
XX protein.
XX
XX PS Claim 1; Page 23; 29pp; English.
XX
XX The present sequence is the yfjK amino acid excretion protein from
XX
XX CC Escherichia coli. This protein is involved in the production of amino
XX
XX CC acids, and an increase in its expression leads to an increased
XX
XX CC accumulation of amino acids in the cell. In this case, an increase in
XX
XX CC threonine, histidine, glutamic acid and proline is achieved if multiple
XX
XX CC copies of its gene are transfected into a bacterium. The bacterium used
XX
XX CC is E. coli
XX
XX SQ Sequence 195 AA;

Query Match 99.7%; Score 993; DB 3; Length 195;
Best Local Similarity 99.5%; Pred. No. 3,1e-109;
Matches 194; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VTPTLLSAFWTYTLITAMTPGPNNTILASATSHGFRQSTRVLAGMSLGLIYVLLCAGI 60
   |||||
Db 1 VTPTLLSAFWTYTLITAMTPGPNNTILASATSHGFRQSTRVLAGMSLGLIYVLLCAGI 60

```

```

QY 61 SFSILAVIDPAVAHLLSWAGAAYIYVWLAKIATSPTEKEDGIQAKPISFWASFPALQFVNVKI 120
   |||||
Db 61 SFSILAVIDPAVAHLLSWAGAAYIYVWLAKIATSPTEKEDGIQAKPISFWASFPALQFVNVKI 120

QY 121 ILVGTALSTFVLPTQOALSWMVGVSVLLAMIGTFGNVCWALAGHLFORLFRQYGRQNT 180
   |||||
Db 121 ILVGTALSTFVLPTQOALSWMVGVSVLLAMIGTFGNVCWALAGHLFORLFRQYGRQNT 180

QY 181 VIALLLVYCAVRIFY 195
   |||||
Db 181 VIALLLVYCAVRIFY 195

RESULT 3
ABO66144 standard; protein; 205 AA.
ID ABO66144
AC ABO66144;
XX
XX 29-JUL-2004 (first entry)
XX
XX Klebsiella pneumoniae polypeptide seqid 12661.
XX
XX Recombinant expression vector; transcription regulatory element;
XX
XX Klebsiella pneumoniae protein; antibacterial; Vaccine.
XX
XX Klebsiella pneumoniae.
XX
XX US6610836-B1.
XX
XX 26-AUG-2003.
XX
XX 27-JAN-2000; 2000US-00489039.
XX
XX 29-JAN-1999; 99US-0117747P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton GL, Osborne M;
XX
XX WPI; 2003-895346/82.
XX
XX N-PSDB; ACH99695.
XX
XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
XX
XX PT preparing a vaccine composition against Klebsiella pneumoniae.
XX
XX PS Disclosure, SEQ ID NO 12661; 932pp; English.
XX
XX The invention describes a new isolated nucleic acid encoding a Klebsiella
XX
XX CC pneumoniae polypeptide. Also described are: a recombinant expression
XX
XX CC vector comprising the nucleic acid, operably linked to a transcription
XX
XX CC regulatory element; and a cell comprising the recombinant expression
XX
XX CC vector. The nucleic acid is useful for preparing a vaccine composition
XX
XX CC against Klebsiella pneumoniae. This is the amino acid sequence of a
XX
XX CC Klebsiella pneumoniae polypeptide of the invention
XX
XX SQ Sequence 205 AA;

Query Match 68.7%; Score 684; DB 7; Length 205;
Best Local Similarity 68.4%; Pred. No. 1,8e-72;
Matches 132; Conservative 21; Mismatches 40; Indels 0; Gaps 0;

QY 1 VTPTLLSAFWTYTLITAMTPGPNNTILASATSHGFRQSTRVLAGMSLGLIYVLLCAGI 60
   |||||
Db 10 VTPTLLSAFWTYTLITAMTPGPNNTILASATSHGFRQSTRVLAGMSLGLIYVLLCAGI 69

QY 61 SFSILAVIDPAVAHLLSWAGAAYIYVWLAKIATSPTEKEDGIQAKPISFWASFPALQFVNVKI 120
   |||||
Db 70 TFSILVEIDSRFTVLVIGIGAGAYIYVWLAKIATSPTEKEDGIQAKPISFWASFPALQFVNVKI 129

QY 121 ILVGTALSTFVLPTQOALSWMVGVSVLLAMIGTFGNVCWALAGHLFORLFRQYGRQNT 180
   |||||
Db 130 ILVGTALSTFVLPTREPVWLISVSLILAAIGALGNTLFWALAGHLFORLFRQYGRQNT 189

```


	Matches	71;	Conservative	39;	Mismatches	67;	Indels	15;	Gaps	8
Oy		5	LLSAFWYTLITAMPGPNNILIASATSHGROSTRVLAGSLGFLIVMLICAGISPSL	64						
Dd		51	LLLAFLSLPAFVTSVPSPGNNTLLASGVNFGVRSRIPHLIGSCGFPIVM--AVGFL	107						
Oy		65	AVI---DPAAHLLSMAGAAYIVMLAMKIATSPKEDCLQAKPISFPASPALQFVNWK	119						
Dd		108	GTVEFAFYPLVTLYILEYGAAYLLIAMKATISGPAQNPEEGKRLPYSGAAAFQWVNRK	167						
Oy		120	IILYEVTLSTFEVLPFOQLSWVGVSYLAMIQTFCGNVCMALAGHLPFLFR-QYG	175						
Dd		168	AMWAHVGAISTYT-FLGGFTNVVVVISYVFALINAPTIQ-IWARGSMTLRVADPRWL	224						
Oy		176	FQNLTVALLLV	187						
Dd		225	RVFNVGMALLLV	236						
RESULT 6										
ID	ABM70439		ABM70439 standard; protein; 203 AA.							
XX	ABM70439;									
AC										
DX	20-NOV-2003	(first entry)								
DT										
DE	Photorhabdus luminescens	protein sequence #3536.								
KW	Antibacterial, fungicide, insecticide, polymorphism; genetic analysis;									
KM	detection; food; gene expression; plant; animal; microorganism; toxin;									
KW	antibiotic; biopesticide; virulence factor; disease model; plague;									
KW	whooping cough.									
OS	Photorhabdus luminescens.									
XX										
FN	WO200294867-A2.									
PD	28-NOV-2002.									
PF	07-FEB-2002; 2002MO-IB003040.									
XX										
PR	07-FEB-2001; 2001FR-00001659.									
XX										
PA	(INSP) INST PASTEUR.									
PI	(CNRS CENT NAT RECH SCI.									
P1	Duchaud E, Tcourit S, Glaeser P, Frangeul L, Kunst F, Danchin A;									
P1	Buchlesser C;									
DR	WPI; 2003-148459/14.									
PT										
XX										
PS	Genomic sequence of Photorhabdus luminescens and encoded polypeptides,									
XX	useful e.g. as therapeutic antimicrobials and agricultural pesticides.									
CC	Claim 2; SEQ ID NO 3536; 1205bp; French.									
CC	The invention relates to the isolation of genes and their encoded									
CC	proteins from Photorhabdus luminescens. The isolated sequences are									
CC	sources of probes and primers for detecting the genome of P. luminescens									
CC	and related species; to study polymorphisms; for gene analysis and for									
CC	detection/amplification of the genes. Antibodies (Ab) raised against the									
CC	polypeptides encoded by the genes are used for detection/identification									
CC	of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that									
CC	carry a gene-containing vector are used to select compounds that									
CC	modulate, regulate, induce or inhibit expression of the genes in plants,									
CC	animals or microorganisms other than P. luminescens and are able to alter									
CC	response or sensitivity to toxins and antibiotics produced by P.									
CC	luminescens. Cells transformed to express the genes are useful for									
CC	recombinant production of the proteins, particularly toxins and									
CC	antibacterials useful as insecticides, bactericides and fungicides. The									
CC	genes, proteins, vectors containing the genes and Ab are also useful									
CC	therapeutically (to treat microbial infection by bacteria or fungi that									

CC	are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC	biopesticides. Other uses of the genes and the proteins are as virulence
CC	factors and for identifying targets of human diseases for which P.
CC	luminescens is a model (particularly plague and whooping cough). This
xx	sequence represents one of the isolated P. luminescens proteins
SQ	Sequence 203 AA;
Query Match	25.7%; Score 256; DB 6; Length 203;
Best Local Similarity	30.8%; Pred. No. 1.3e-21;
Matches	62; Conservative 47; Mismatches 84; Indels 8; Gaps 4
QY	1 VPTPLLSAFWYTTLTMTTPGPNNTLLASSATSHSGRQSTRVLAMSGELVIMLLCA-G 59
DB	1 MTASLIFSLTYPLFISSTIPGNMMLTRSSGANFGFIRSLGLFVGVLDMQTLLAARG 60
QY	60 ISFSLAVIDPPAAVHLISWAGAYIYWLAKVIKTSPTK---EDGLQAKDISEFWASEPALOF 115
DB	61 IALALLVPRAIHVMKTIAGSLYLMLMAKKTITASTRYLDTDSDAVIKRVRYQGWLDP 119
QY	116 VNVKIILGVTAALSTFVLPRQDALSMVVGVSVLLAMIGTFGNVCVALACHLFQRLEFRQYG 175
DB	120 LNPXAMLMGLGAVGSYSIAGDHLYSIIVTSIVMFNINLCGMVWAGFSSLIGTLRLSRN 179
QY	176 R-QNLIVALLLVTCVAVRIF 194
DB	180 AMFLFNITMGILTATCVLPIM 200

	RESULT 7
AB063925	
ID	AB063925 standard; protein; 211 AA.
XX	
AC	AB063925;
XX	
DT	29-JUL-2004 (first entry)
XX	
DE	Klebsiella pneumoniae polypeptide seqid 10442.
XX	
KW	Recombinant expression vector; transcription regulatory element;
KW	Klebsiella pneumoniae protein; antibacterial; Vaccine.
XX	
OS	Klebsiella pneumoniae.
XX	
PN	US6610836-B1.
PD	
XX	26-AUG-2003.
PF	
XX	27-JAN-2000; 2000US-00489039.
PR	
XX	29-JAN-1999; 99US-0117747P.
PA	(GENO-) GENOME THERAPEUTICS CORP.
PI	
XX	Breton GL, Osborne M;
DR	WPI; 2003-895346/82.
DR	N-PsDB; ACH97476.
XX	
PT	New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for preparing a vaccine composition against Klebsiella pneumoniae.
XX	
PS	Disclosure; SEQ ID NO 10442; 932pp; English.
XX	
CC	The invention describes a new isolated nucleic acid encoding a Klebsiella pneumoniae polypeptide. Also described are: a recombinant expression vector comprising the nucleic acid, operably linked to a transcription regulatory element; and a cell comprising the recombinant expression vector. The nucleic acid is useful for preparing a vaccine composition against Klebsiella pneumoniae. This is the amino acid sequence of a Klebsiella pneumoniae polypeptide of the invention
SQ	Sequence 211 AA;

Query Match	25.2%	Score 230.5	DB 7	Length 211
Best Local Similarity	32.4%	Pred. No. 6e-21		
Matches	61	Conservative 42	Mismatches 78	Indels 7
				Gaps 5
Qy	1LSAFWTYTLITAMPGPNNILASATSHGROSTRVLAGWSLGLVIMLLCAGISFSL	64		
Db	17 ILIAITIAFPTSTIRPGNNMLLSGLNGFKRSLPHYLGISIGFAL-MYIAVETGICR	75		
Qy	65 AVIDPAVHL-LSWGAAYIVWLANKIATS--PYKEDGLQAKPISFMWSPALQFNVXII	121		
Db	76 VFENMPQYVIALRICGATYTLIYLAWNITDALPMSKISSSSARPSFWEAAGQWNPXAM	135		
Qy	122 LYGVALSTFPLPQGLQALSWVGVSVLAMICTFGNVCMALAGHLFQRLFR--QYGRQLN	179		
Db	136 IMATGATITTY-LPQGENVRSVILFSLIPASINAPSVSIWTFGATLRHMLNDIKYRIEN	194		
Qy	180 IVLALLLV 187			
Db	195 VGMAMLLL 202			
RESULT 8				
ADFO4482				
ID	ADFO4482	standard; protein; 212 AA.		
XX	ADFO4482;			
AC				
XX	12-FEB-2004 (first entry)			
DT				
XX	Bacterial polypeptide #595.			
DE				
XX	Proteus mirabilis infection; bacterial infection; antibacterial;			
KM	immunoculant.			
XX				
XX	Proteus mirabilis.			
OS				
XX	US6605709-B1.			
PN				
XX	12-AUG-2003.			
PD				
XX	05-APR-2000; 2000US-00543681.			
PF				
XX	09-APR-1999; 99US-0128706P.			
PR				
XX	(GENO-) GENOME THERAPEUTICS CORP.			
PA				
Breton GL;				
PI				
XX	WPI: 2003-895291/82.			
DR	N-PSDB; ADF00310.			
XX				
XX				
PS	Disclosure; SEQ ID NO 4767; 870pp; English.			
XX				
XX	The invention relates to new Proteus mirabilis polypeptides and			
CC	polynucleotides. The invention also relates to antibodies against the			
CC	polypeptides, methods for producing the polypeptides, a method of			
CC	generating vaccines for immunising an individual against P. mirabilis, a			
CC	method for evaluating a compound for the ability to bind a P. mirabilis			
CC	polypeptide and a method for screening test compounds for anti-bacterial			
CC	activity. The polypeptides and polynucleotides are useful for molecular			
CC	targets for diagnosing, preventing and treating pathological conditions			
CC	resulting from bacterial infection, as reagents for diagnosis of			
CC	bacterial diseases, as components of antibacterial vaccines, as targets			
CC	for antibacterial drugs or as bio-control agents for plants. This			
CC	sequence represents a Proteus mirabilis polypeptide of the invention.			
Q0	Sequence 212 AA;			

Query Match	23.3%	Score 232.5	DB 7	Length 212
Best Local Similarity	30.2%	Pred. No. 8.3e-19		
Matches	57	Conservative 45	Mismatches 80	Indels 7
				Gaps 6
QY	12	YTLTAMPGNNITLSSATSHQPROSTRVLAMSGFLIVMLQGISF-SLAVIDPA	70	
Db	25	FSPTSTPGNNITMLASGINFELKRTMPHAIVSGF-FVMTLAVIGIGALIKSP	83	
QY	71	AVHLTSMGAAYIVMLAKTIATSPTEK--DGLQAKPISFMAFALQFVNVIILYGTAL	128	
Db	84	IYNILKTYIGALYMLAMKTAISVSQNSKQSPPLTEBALFQVIRNKSMMALSGI	143	
QY	129	STFVLPQTQALSMVVGVSVLMIATGTGNCWALAGH-LFORLFR-QYGRQNLIVALL	186	
Db	144	TLVYTSPOVPIYSMLL-VAIIFTLINFPCVAIMATFGSHLRRLKNPKILKLFNIMGGL	202	
QY	187	YVCAYRIPY	195	
Db	203	ALSAISVLF	211	
RESULT 9				
ADP06423	ID	ADP06423 standard; protein, 211 AA.		
AC		ADP06423;		
XX				
XX		12-FEB-2004 (first entry)		
XX				
DE		Bacterial polypeptide #2536.		
XX				
XX		Proteus mirabilis infection; bacterial infection; antibacterial;		
KM		immunostimulant.		
XX				
OS		Proteus mirabilis.		
XX				
PN		US6605709-B1.		
XX				
PD		12-AUG-2003.		
XX				
PF		05-APR-2000; 2000US-00543681.		
XX				
PR		09-APR-1999; 99US-0128706P.		
XX				
PA		(GENO-) GENOME THERAPEUTICS CORP.		
XX				
PI		Breton GL;		
XX				
DR		WPI: 2003-895291/82.		
XX				
DR		N-PSDB; ADP02251.		
XX				
PT		New Proteus mirabilis polypeptides and polynucleotides, useful as		
PT		reagents for diagnosis of bacterial disease, as components of		
PT		antibacterial vaccines, as targets for antibacterial drugs, or as		
PT		biocontrol agents for plants.		
XX				
PS		Disclosure; SEQ ID NO 6708; 870bp; English.		
XX				
CC		The invention relates to new Proteus mirabilis polypeptides and		
CC		polynucleotides. The invention also relates to antibodies against the		
CC		polypeptides, methods for producing the polypeptides, a method of		
CC		generating vaccines for immunising an individual against P. mirabilis, a		
CC		method for evaluating a compound for the ability to bind a P. mirabilis		
CC		polypeptide and a method for screening test compounds for anti-bacterial		
CC		activity. The polypeptides and polynucleotides are useful as molecular		
CC		targets for diagnosing, preventing and treating pathological conditions		
CC		resulting from bacterial infection, as reagents for diagnosis of		
CC		bacterial diseases, as components of antibacterial vaccines, as targets		
CC		for antibacterial drugs or as bio-control agents for plants. This		
CC		sequence represents a Proteus mirabilis polypeptide of the invention.		
XX				
SQ		Sequence 211 AA;		

Matches	32	Conservative	46	Mismatches	79	Indels	15	Gaps	8
QY	6	LSAFWYTLITMTPEGNNTIALSSATSHGFROSTRVLAMSGILVLLCAGISFSLA	65						
DB	41	LTATSLPAIVASTIPGPTNFILSLSSHYKISKTLPIVLGSCIGALLVLV-VGIGLST	99						
QY	66	VTD-PAAVHLISWGAIVYIWLAMKIT-SP--TKEGLOAKPSFWASPALQFVANKII	121						
DB	100	ILAAVPIQKIMASGILMTVLAMKRLVYNPVSILENEQYPPGFGFAEPLMOAINKTW	159						
QY	122	LYGTALSTFVLPQTQALSWVGVSVLAMIGTFGNVC--WMALAGHLFQRLFRQYGRQL	178						
DB	160	MMAFAVIVSVYI--KQGGDILIVNSIISLCIFLIAFPCLYIMALVGRSLRSLSK-PKH	215						
QY	179	NI---VLALLV	187						
DB	216	NIFFKIMAILL	227						
RESULT 11									
ID	ADA35872								
AD	ADA35872	standard; protein; 241 AA.							
AC	ADA35872;								
XX	20-NOV-2003	(first entry)							
DT	Acinetobacter baumannii	protein #3033.							
DE	Acinetobacter baumannii	bacterial disease; antibacterial; vaccine;							
XX	plant biocontrol agent.								
XX	Acinetobacter baumannii.								
OS	US6562958-B1.								
PN	13-MAY-2003.								
XX	04-JUN-1999;	99US-00328352.							
XX	09-JUN-1998;	98US-0088701P.							
PR	(GENO-)	GENOME THERAPEUTICS CORP.							
PA	Breton G, Bush D;								
XX	WPI; 2003-576092/54.								
DR	N-PSDB; ADA31746.								
XX	New Acinetobacter baumannii	proteins and nucleic acid, useful as reagents							
PT	for diagnosing a bacterial disease,	as components of antibacterial							
PT	vaccines, as targets for antibacterial	drugs, or as biocontrol agents for							
PI	plants.								
XX	Example; SEQ ID NO 7159;	328bp; English.							
PS	The invention relates to isolated	Acinetobacter baumannii nucleic acids.							
XX	The A. baumannii nucleic acid	and polypeptides are useful as reagents							
CC	for diagnosing a bacterial disease,	as components of antibacterial							
CC	vaccines, as targets for antibacterial	drugs, to detect the presence of							
CC	A. baumannii and other Acinetobacter	species in a sample, in screening							
CC	compounds for the ability to interfere	with the A. baumannii life cycle							
CC	or to inhibit A. baumannii infection,	and as biocontrol agents for							
CC	plants. The present sequence	represents the amino acid sequence of an A.							
CC	baumanni protein.								
XX	Sequence 241 AA;								
QY	Query Match	17.9%; Score 178; DB 6; Length 241;							
QY	Best Local Similarity	24.6%; Pred. No. 3e-12;							

Db 34 AFGICLAWLTGPGNNIYILSRISGCKIAGFISLGGVAVGFYFW-LCAASGITLTV 92
QY 67 IDPAVHLISWAGAAIYVMTAKIATISPTKEDGLQAKPIS-----FMASPALQFVNVK 119
Db 93 AVPAAYDTIRIAGAMYLMLAMK-ALRPNAPAFENVDLADSPILKFLMGFLTNLNPX 151
QY 120 IILYGVATLSPVLPQTQALSWVVGVSVLAMIGTFGNV 158
Db 152 IAIWYLSLPPQFIHPQGS---IIAQSIQLGTIOIFVSV 187

RESULT 14

ABO61549

ID ABO61549 standard; protein: 214 AA.

XX ABO61549;

XX 29-JUL-2004 (first entry)

XX Klebsiella pneumoniae polypeptide seqid 8066.

XX Recombinant expression vector; transcription regulatory element;

KM Klebsiella pneumoniae protein; antibacterial; Vaccine.

XX Klebsiella pneumoniae.

XX US6610836-B1.

XX 26-AUG-2003.

XX 27-JAN-2000; 2000US-00489039.

XX 29-JAN-1999; 99US-0117747P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Breton GL, Osborne M;

XX WPI; 2003-895346/82.

XX N-PSDB; ACH95100.

PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
PT preparing a vaccine composition against Klebsiella pneumoniae.

XX Disclosure; SEQ ID NO 8066; 932pp; English.

XX The invention describes a new isolated nucleic acid encoding a Klebsiella

CC pneumoniae polypeptide. Also described are: a recombinant expression

CC vector comprising the nucleic acid, operably linked to a transcription

CC regulatory element; and a cell comprising the recombinant expression

CC vector. The nucleic acid is useful for preparing a vaccine composition

CC against Klebsiella pneumoniae. This is the amino acid sequence of a

CC Klebsiella pneumoniae polypeptide of the invention

XX Sequence 214 AA;

Query Match 16.3%; Score 162; DB 7; Length 214;

Best Local Similarity 26.5%; Pred. No. 2e-10;

Matches 54; Conservative 42; Mismatches 82; Indels 26; Gaps 10;

QY 5 LLSAFMTY-----TLTAMTPGPNNIILASATSHGFROSTRVLAGMSLGFLLVMLCA- 58

Db 8 LMTLEWMPAYILTJITLISLSRSGAINTMTSTINHGVRGAASAGIQGTGLAHIVGV 67

QY 59 --GISFSIAVIDPAVHLISWAGAAIYVMTA--WKIATSPTEKD--GLQAKPISFWASF 111

Db 68 GLGTLFSSRYL---AFEVTLKWAAGAAIYILGIIQWRRAAGIDLMTLAKAQTRGKLFQRAV 124

QY 112 ALQGVNKKIILYGTALSTFLP-OTQALSWV-GVSV---LLAMIG--TFGN--VCWA 161

Db 125 FVNLTNPKSIVFLAALFPQFILPHQPOVMQYIVLGVTTIVVDIIVIGVATLAQRISAMI 184

QY 162 LAGHLFQRLFRQYGRQTLNIVALL 185
Db 185 KGPXOMKALNKVFGSLFMLVGLL 208

RESULT 15

ABU32233

ID ABU32233 standard; protein: 206 AA.

XX ABU32233;

XX 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #1760.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Klebsiella pneumoniae.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0326599P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreych RA, Xu HH;

XX WPI; 2003-029926/02.

XX N-PSDB; ACA36103.

PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 60157; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation; (7) identifying a gene in an operon required for

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this

Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 10, 2005, 06:45:53 ; Search time 3392 Seconds
(without alignments)
10713.860 Million cell updates/sec

Title: US-10-620-487-1

Perfect score: 750
Sequence: 1 gatccataaccacaaccta.....ttagtaactgtcacttaag 750

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb ba: *
2: gb hlg: *
3: gb_in: *
4: gb_com: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_por: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	750	100.0	750	6	CQ759947
2	750	100.0	10272	1	ECORABAB
3	750	100.0	16446	1	D90886 E.coli geno
4	750	100.0	29254	1	ECOK12R11
5	750	100.0	110000	1	U000096_27
6	745.2	99.4	12322	1	AE015279
7	745.2	99.4	290380	1	AE016987
8	738.8	98.5	300099	1	AE016764
9	687.6	91.7	12337	1	AE005488
10	687.6	91.7	270365	1	AP002562
11	623.4	83.1	194171	2	AC126452
12	588	78.4	588	6	E49391
13	588	78.4	588	6	AX030083
14	461.2	61.5	145050	1	AL627275
15	461.2	61.5	300431	1	AE016835
16	458	61.1	21722	1	AE008820
17	265.8	35.4	618	6	AR388761
18	224.4	29.9	23738	2	AC020828
19	156	20.8	372	6	AR388768

20	100	13.3	100	6	AX999538	AX999538 Sequence
21	100	13.3	100	6	AX999539	AX999539 Sequence
22	100	13.3	100	6	AX999540	AX999540 Sequence
23	72.6	9.7	110000	1	AE017333_05	Continuation (6 of
24	72.6	9.7	110000	1	CP000002_05	Continuation (6 of
25	70.4	9.4	288404	1	AE017272	AE017272 Bacillus
26	63.8	8.5	302451	1	AE016812	AE016812 Vibrio vu
27	55.2	7.4	110000	1	AE017225_22	Continuation (23 o
28	55.2	7.4	110000	1	AE017334_22	Continuation (23 o
29	55.2	7.4	110000	1	AE017355_23	Continuation (24 o
30	55.2	7.4	110000	1	CP000001_23	Continuation (24 o
31	55.2	7.4	293264	1	AE017031	AE017031 Bacillus
32	51	6.8	249500	1	AP005344	AP005344 Vibrio vu
33	50.4	6.7	573	6	AX432019	AX432019 Sequence
34	49.6	6.6	300600	1	AP005084	AP005084 Vibrio pa
35	49	6.5	110000	1	BX950851_45	Continuation (46 o
36	48.8	6.5	344805	1	BX640434	BX640434 Bordetell
37	48.8	6.5	346287	1	BX640450	BX640450 Bordetell
38	47.8	6.4	304262	1	AE017005	AE017005 Bacillus
39	47.2	6.3	13848	1	AE004249	AE004249 Vibrio ch
40	47.2	6.3	348171	1	BX640412	BX640412 Bordetell
41	44.4	5.9	310029	1	AE016868	AE016868 Zebrafish
42	42.8	5.7	142350	5	BX088596	BX088596 Zebrafish
43	42	5.6	669	6	AR319795	AR319795 Pseudom
44	41.4	5.5	301214	1	AE016786	AE016786 Pseudom
45	41	5.5	618	12	AY657639	AY657639 Synthetic

ALIGNMENTS

RESULT 1
LOCUS CQ759947
DEFINITION Sequence 1 from Patent EP1382684.
ACCESSION CQ759947
VERSION CQ759947.1 GI:44903612
KEYWORDS
SOURCE
ORGANISM Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE
AUTHORS Maier,T.
TITLE Process for fermentative production of amino acids and amino acid derivatives of the phosphoglycerate family
JOURNAL Patent: EP 1382684-A 1 21-JAN-2004;
Consortium fuer elektrochemische Industrie GmbH (DE)

FEATURES
source
1..750
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
110..697
/note="unassigned protein product"

CDS

/codon_start=1
/translation="MTPTLSAFWYTLITAMTPGPNILALSSATSHGPROSTRVLA
GMSIGPLIVMLCAGISPSFLAVIDPAVHLISWAGAAVYVIMAKRTASPTEDGLOA
KPISTFASPALQFVANKIITLYGTALSTFVLDTOTGLSNVGVSVILAMIGTRGNVCM
ALAHGLFRLFRYGRQNLIVLALLVYCAVRIY"

ORIGIN

Query Match 100.0%; Score 750; DB 6; Length 750;
Best local Similarity 100.0%; Pred. No. 1.7e-224; Indels 0; Gaps 0;
Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCCATAACCCCAACCTATCGAATAATTCGAATCTAGAAATTAATAAAATTCATTATTTT 60
1 GATCCATAACCCCAACCTATCGAATAATTCGAATCTAGAAATTAATAAAATTCATTATTTT 60

Qy 61 TTAATGTCGCTGCGGTAAGTCTTACCAAAAAGAGAGATTAACAAGTGAACCCGAC 120
 |||||
 Db 61 TTAATGTCGCTGCGGTAAGTCTTACCAAAAAGAGAGATTAACAAGTGAACCCGAC 120
 |||||
 Qy 121 CTTTAAAGTCTTTTGGACTTACCCCTGATTACCGCTATGACCGCAGACCGAACA 180
 |||||
 Db 121 CTTTAAAGTCTTTTGGACTTACCCCTGATTACCGCTATGACCGCAGACCGAACA 180
 |||||
 Qy 181 TATTCGCGCCCTTACGCTGCTACGAGCATGATTTGTCMAAGTACCCGCGCTGGC 240
 |||||
 Db 181 TATTCGCGCCCTTACGCTGCTACGAGCATGATTTGTCMAAGTACCCGCGCTGGC 240
 |||||
 Qy 241 AGGATGATGCTGGGATTTTGTATGATGATTAAGTGTGCGGCGCATTTTCATTTTCACT 300
 |||||
 Db 241 AGGATGATGCTGGGATTTTGTATGATGATTAAGTGTGCGGCGCATTTTCATTTTCACT 300
 |||||
 Qy 301 GGCAGTATGACCCGCGAGCGGTAACCTTTTGAAGTTGGCGGGCGGCGCATATTTGT 360
 |||||
 Db 301 GGCAGTATGACCCGCGAGCGGTAACCTTTTGAAGTTGGCGGGCGGCGCATATTTGT 360
 |||||
 Qy 361 CTGGCTGGCGGTAAGGCAATGCGCACAGCCCAAGAAAGAGACGATTCAGGCAAAACC 420
 |||||
 Db 361 CTGGCTGGCGGTAAGGCAATGCGCACAGCCCAAGAAAGAGACGATTCAGGCAAAACC 420
 |||||
 Qy 421 AATCAGCTTTTGGCGCAGCTTTTGTGAGTTTGTACGCTCAAAATCATTTTGTATCGG 480
 |||||
 Db 421 AATCAGCTTTTGGCGCAGCTTTTGTGAGTTTGTACGCTCAAAATCATTTTGTATCGG 480
 |||||
 Qy 481 TGTATCGGCACTGTGACGTTTGTCTGCGGCAAAACACAGCGCTTAAGCTGGATTTGG 540
 |||||
 Db 481 TGTATCGGCACTGTGACGTTTGTCTGCGGCAAAACACAGCGCTTAAGCTGGATTTGG 540
 |||||
 Qy 541 CGTCAGGCTTTTGTGCGGATGATTTGGACGTTTGGCAATGTGTGCTGGCGCTGGCGGG 600
 |||||
 Db 541 CGTCAGGCTTTTGTGCGGATGATTTGGACGTTTGGCAATGTGTGCTGGCGCTGGCGGG 600
 |||||
 Qy 601 GCATCTGTTTCAGGATTTTTCGCCAGATATGTGTCGCCAGTTAAATTCGTGTTGCCCT 660
 |||||
 Db 601 GCATCTGTTTCAGGATTTTTCGCCAGATATGTGTCGCCAGTTAAATTCGTGTTGCCCT 660
 |||||
 Qy 661 GTTCTGTGCTATTTGCGCGGATGAGCATTTTCTATTAACGAAAAAAGCGAAGAGGTGCG 720
 |||||
 Db 661 GTTCTGTGCTATTTGCGCGGATGAGCATTTTCTATTAACGAAAAAAGCGAAGAGGTGCG 720
 |||||
 Qy 721 CCTCTCGCGTTAGTAATTCGCTACTTAAG 750
 |||||
 Db 721 CCTCTCGCGTTAGTAATTCGCTACTTAAG 750
 |||||

RESULT 2
 ECOMBAB 10272 bp DNA linear BCT 14-APR-2000
 LOCUS E.coli genes for L-aspartate oxidase, putative ATP dependent RNA
 DEFINITION helicase, and uracil DNA glycosylase.
 ACCESSION D13169
 VERSION D13169.1 GI:285775
 KEYWORDS ATP dependent RNA helicase; L-aspartate oxidase; uracil DNA
 glycosylase.
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE
 AUTHORS 1 (bases 1 to 10272)
 TITLE Nashimoto, H., Mura, A., Saito, H. and Uchida, H.
 Suppressor of temperature-sensitive mutations in a ribosomal
 protein gene, rpsL (S12), of Escherichia coli K12
 JOURNAL Mol. Gen. Genet. 199 (3), 381-387 (1985)
 MEDLINE 85295454
 PUBMED 3162077
 AUTHORS 2 (sites)
 REFERENCE Varshney, U., Hutcheon, T. and van de Sande, J.H.
 TITLE Sequence analysis, expression, and conservation of Escherichia coli
 uracil DNA glycosylase and its gene (ung)

JOURNAL J. Biol. Chem. 263 (16), 7776-7784 (1988)
 MEDLINE 88227981
 PUBMED 2836397
 REFERENCE 3 (sites)
 AUTHORS Plachmann, R., Kunz, N., Seifert, J., Gutlich, M., Wientjes, F.J.,
 Laufer, A. and Gassen, H.G.
 TITLE Molecular biology of pyridine nucleotide biosynthesis in
 Escherichia coli. Cloning and characterization of quinolinate
 synthase genes nadA and nadB
 JOURNAL Eur. J. Biochem. 175 (2), 221-228 (1988)
 MEDLINE 88296484
 PUBMED 2841129
 REFERENCE 4 (sites)
 AUTHORS Nishi, K., Morel-Deville, F., Hershey, J.W., Leighton, T. and
 Schuster, J.
 TITLE An eIF-4A-like protein is a suppressor of an Escherichia coli
 mutant defective in 50S ribosomal subunit assembly
 JOURNAL Nature 336 (6198), 496-498 (1988)
 MEDLINE 89057150
 PUBMED 2461520
 REFERENCE 5 (sites)
 AUTHORS Nashimoto, H.
 TITLE Non-ribosomal proteins affecting the assembly of ribosomes in
 Escherichia coli
 JOURNAL (in) Niehaus, K.H. (Ed.):
 THE TRANSLATIONAL APPARATUS: 185-195;
 Plenum Publishing Corp., New York, U.S.A. (1993)
 REFERENCE 6 (sites)
 AUTHORS Lonetto, M.A., Brown, K.L., Rudd, K.E. and Butner, M.J.
 TITLE Analysis of the Streptomyces coelicolor sigE gene reveals the
 existence of a subfamily of eubacterial RNA polymerase sigma
 factors involved in the regulation of extracytoplasmic functions
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (16), 7573-7577 (1994)
 MEDLINE 94329558
 PUBMED 8052622
 REFERENCE 7 (bases 1 to 10272)
 AUTHORS Nashimoto, H.
 TITLE Direct Submision
 JOURNAL Submitted (02-SEP-1992) Hiroko Nashimoto, Teikyo University,
 Department of Biosciences, Toyosatodai, Utsunomiya, Tochigi 320,
 Japan (Tel:028-627-7181, Fax:028-627-7181)
 COMMENT Submitted (02-SEP-1992) to DDBJ by:
 Hiroko Nashimoto
 Dept. of Biosciences
 Teikyo University
 Toyosatodai, Utsunomiya
 Tochigi 320
 Japan
 Phone: 0286-27-7181
 Fax: 0286-27-7187.
 FEATURES
 source 1..10272
 Location/Qualifiers
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="K-12"
 /db_xref="taxon:562"
 1642..3264
 /gene="nadB"
 1642..3264
 /gene="nadB"
 /citation=[3]
 /codon_start=1
 /transl_table=11
 /product="L-aspartate oxidase"
 /protein_id="BAB02446.1"
 /_id_xref="GI:285775"
 /_translacion="NNTLPEHSCDVLIISSGAAGSLALRLADQHOVILSKPYTEG
 STFYAQGIAAVDETDSDSHVEDTLIAGICDRHAYEPVANSARSCVWLIDQGV
 LFDTHIQPNEBSYHLTRGSGSHRIILHAADATGREGVETTLVSKALNPINIRLERT
 NAYDLIVSDKIGIGRTRVGVAMVMNRKETEYETCHAKAVLATGASVYQYTTTPD
 ISSGDSIINAMRAGCRVNLERNQPHPTALYHPQRNPLLTALRREGAYLRPDGTR
 FMDPDERSELARDIVARALIDEMKRLADCMFLDISKPADFLIQHPMIYKILG
 LGIDILQEPVPIVPAHAYTCGVMVDHGRITVEGLYALGEVSYTELHGANNRANSL

TITLE
Itoh, T., Kimura, S., Kitagawa, M., Makino, K., Miki, T., Mitsuhashi, N., Mizobuchi, K., Mori, H., Nakade, S., Nakamura, Y., Nishimoto, H., Oshima, T., Oyama, S., Saito, N., Sampei, G., Sato, Y., Takemoto, K., Sivasubramaniam, S., Tagami, H., Takahashi, H., Takeda, J., Takemoto, K., Uehara, K., Wada, C., Yamagata, S. and Horiuchi, T.
Construction of a contiguous 874-kb sequence of the *Escherichia coli* -K12 genome corresponding to 50.0-68.8 min on the linkage map and analysis of its sequence features

JOURNAL
DNA Res. 4 (2), 91-113 (1997)

MEDLINE
97349980

PUBMED
9205837

REFERENCE
2 (sites)

AUTHORS
Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y. and Yano, M.
The systematic sequencing of the *Escherichia coli* genome in Japan
Unpublished
3 (bases 1 to 16446)

TITLE
Mori, H.
Direct Submission
Submitted (23-JAN-1997) Hirotsada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info., 8916-5 Takayama, Ikoma, Nara 630-01, Japan
(E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)
Collaboration Information:
Project:

COMMENT
The Japan E. coli genome DNA sequencing project
Group:
The Japan E. coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y. and Yano, M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotsada Mori
Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan
E-mail: hmori@gtc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http:bsw3.aist-nara.ac.jp.
Location/Qualifiers
1. 16446
/organism="Escherichia coli"
/mol type="genomic DNA"
/strain="K12"
/db_xref="taxon:562"
/map="58.4 min"
/clone="Kohara clone #436"
/clone_1lb="Kohara lambda miniset library"
/note="Nucleotide position 2709472-2725917 from the initiation site of *Pha* (0 min.).-This clone is from Kohara lambda miniset library"
complement(1. 623)
/gene="yfiE"
complement(<1. 623)
/gene="yfiE"
/note="similar to [SwissProt Accession Number P33634]"
/codon_start=1

gene
726. 1313
/gene="yfiK"
726. 1313
/note="similar to [SwissProt Accession Number P38101]"
/codon_start=1
/transl_table=11
/protein_id="BAA16464.1"
/db_xref="GI:1799982"
/translation="MTPTLLSFWTYTTLTMTPPNNIILASATSGFQSTPEVLA
GMSLGFVLMLCAGISFLSALIDPAAHLLSWAGAAITVLMKATISPTREGLQIA
KPSIFWASPALQFVNWKIILXGVTALSTFVLFPQTALSMVGVSVLLMIGTFGNVCW
ALAGHLFQRLFPQYGRQNLIVALLVCAVAFY"
complement(1369. 1752)
/gene="yfiD"
complement(1369. 1752)
/note="similar to [SwissProt Accession Number P33633]"
/codon_start=1
/transl_table=11
/protein_id="BAA16465.1"
/db_xref="GI:1799983"
/translation="MITGIQITKAAANDILNSFWLIDSEKGEARCTVAKAGAEDEV
AVSKGLDIYREVPEVPEVEGQHLNVNLTRETLDEAVKPEKYPQTLITVSG
VAVFNSLTPEEQRDVIARTFTESL"
2057. 2746
/gene="ung"
2057. 2746
/note="similar to [PIR Accession Number A28175]"
/codon_start=1
/transl_table=11
/product="uracil-DNA glycosylase (BC 3.2.2.-)"
/protein_id="BAA16466.1"
/db_xref="GI:1799984"
/translation="MANELTWDHVALEEKQPYPLNTIQTVASERQSGVTYIPQKDV
FNAFRTELDVAVVILGQDPYHGQAGLAFSVPGIAPPSILNMYKLENTIPG
FTRPNHGVSANRQGVLLNTVLTIVRAQASHSLSGWETFTKVI SLINHGSGVV
FILMGSHAKKGAIIIDKQHHVLYKAPHPRLPAHNGFGFCNHFVLANQLQEGRGPPI
DMWFLVPAESE"
complement(2577. 3914)
/gene="yfiF"
complement(2577. 3914)
/note="similar to [SwissProt Accession Number P33635]"
start codon is not identified yet"
/codon_start=1
/transl_table=11
/protein_id="BAA16467.1"
/db_xref="GI:1799985"
/translation="SQALVRNNGEPRLGRLVICALPAALYKESNMDEMKSGKVKMY
VRSDDSDKRTNHPRTGKGGRGKSRADGRRPARDQSQSPRRKMDSPRTYSR
APGDETPERKADHGIGSGSFIDEVLRQRAETRYGNAQCALFKSRNGATTVOQW
IQGVTRPEKALRWMAANKRAYHVVDEALITASGTEHNGVCFLLKSNGETTVQGV
SOAGQDCVIALENESNPNIIGMMRSCHAFGVGVVDGALLLEGAALRTAEGGAE
HNOPTGDNIVNVLTDPFRAGYVVTVTSEOGKPLKTSLPAMKVLVLOEYEGLEDA
ARPNDLRKIDGTGNVAGLNTISVQAPFLCUNGVRIRKPKDSAGNQLCHPAFRKPTH
SLIVLIASSQASARHVVATPDLPAQNGCSQRIHDAPKATDAVLVSHDVAIYL"
4010. 4243
/gene="yfiG"
<4010. 4243
/gene="yfiG"
/note="similar to [SwissProt Accession Number P33636]"
start codon is not identified yet"
/codon_start=1
/transl_table=11

JOURNAL. synthesis genes *nadA* and *nadB*
MEDLINE Eut. J. Biochem. 175 (2), 221-228 (1988)
PUBMED 88296484
REFERENCE 2841129
AUTHORS 6
TITLE Schendel,P.J., Mueller,E., Stubbe,U., Shiau,A. and Smith,J.M.
Formylglycinamide ribonucleotide synthetase from *Escherichia coli*:
cloning, sequencing, overproduction, isolation, and
characterization
JOURNAL. Biochemistry 28 (6), 2459-2471 (1989)
MEDLINE 89274173
PUBMED 2659070
REFERENCE 7
AUTHORS Morrison,P.T., Lovett,S.T., Gilson,L.E. and Kolodner,R.
TITLE Molecular analysis of the *Escherichia coli* *reco* gene
JOURNAL. J. Bacteriol. 171 (7), 3641-3649 (1989)
MEDLINE 89291705
PUBMED 2544549
REFERENCE 8
AUTHORS Poulsen,L.K., Iarsen,N.W., Molin,S. and Andersson,P.
TITLE Analysis of an *Escherichia coli* mutant strain resistant to the
cell-killing function encoded by the *gef* gene family
JOURNAL. Mol. Microbiol. 6 (7), 895-905 (1992)
MEDLINE 92292954
PUBMED 1602968
REFERENCE 9
AUTHORS Nashimoto,H.
TITLE Non-ribosomal proteins affecting the assembly of ribosomes in
JOURNAL. *Escherichia coli*
MEDLINE (in) Nierhaus,K.H. (Ed.) ;
PUBMED THE TRANSLATIONAL APPARATUS: 185-195;
REFERENCE Plenum Publishing Corp., New York, U.S.A. (1993)
AUTHORS 10
TITLE Hiratsu,K., Amemura,M., Nashimoto,H., Shinagawa,H. and Makino,K.
The *ipoE* gene of *Escherichia coli*, which encodes sigma E, is
essential for bacterial growth at high temperature
JOURNAL. J. Bacteriol. 177 (10), 2918-2922 (1995)
MEDLINE 95270615
PUBMED 7751307
REFERENCE 11
AUTHORS Yu,H., Schurr,M.J. and Derecic,V.
TITLE Functional equivalence of *Escherichia coli* sigma E and *Pseudomonas*
aeruginosa AlgU: *E. coli* *ipoE* restores mucoidy and reduces
sensitivity to reactive oxygen intermediates in algu mutants of *P.*
aeruginosa
JOURNAL. J. Bacteriol. 177 (11), 3259-3268 (1995)
MEDLINE 95286510
PUBMED 7768826
REFERENCE 12 (bases 1 to 29254)
AUTHORS Nashimoto,H. and Saito,N.
TITLE Kohara library: 7G4-4A12 contig
JOURNAL. Unpublished
AUTHORS 13 (bases 1 to 29254)
TITLE Nashimoto,H.
JOURNAL. Direct Submission
AUTHORS Submitted (03-SEP-1995) Hiroko Nashimoto, Teiyo University,
MEDLINE Department of Biosciences, Toyosato-ai, Utsunomiya, Tochigi 320,
PUBMED Japan (E-mail:nashimoto@u-u.ac.jp, Tel:01-28-627-7161,
REFERENCE Fax:01-28-627-7181)
FEATURES Location/Qualifiers
SOURCE 1. .29254
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:562"
/clone_lib="Kohara"
/note="Clones 7G4, 4A12"
3406. .4767
/citation=8
/codon_start=1
/transl_table=11
/product="YFED-ECOLI protein similarity"
/protein_id="BA10908.1"

/db_xref="GI:987636"
/translation="MSTHTPLTVNEINCKPGFLDPEYELAKOPADYLGVLKYTVRQNI
SOLFDDJLNGNADILAAGLVNSERVKNVQPGPTVYSVQOLVYKVGVRPRTGLT
ADRLTAPRGHVYVNDLQTLKETKPELSKRVDDKKGSALIMEDVIEGRDYLTAISVA
ISLFGQVHPELVADLIDTDEQPVTFWSPPLDGNVLSAALLDFNEKNEDGTLARIBEK
YLGHGDYVYDTRTFRLRAVDVLPQKPLFKRYAEELDMRLIAAAYDSMDAQAT
SPGVYGMMLTKNTQSLIGTRDPAEQSIGGVAYIQDMSKVPESEVPEMERLWFA
LAAVYNGVAMHMDARALTAKTGNPDSNADVQRPLPSQKPYKSLTGYVARGHAY
AYVENIRKQISLVGTLQEKERQATEAANQLAQDYPVAVLSRMAKNPLFSRPFPSRH
QTI"
3701. .3704
/citation=6
/citation=6
/replace="c"
3753. 3756
/citation=6
/citation=6
/replace="g"
3804. .3807
/citation=6
/replace="g"
4697. .4700
/citation=8
/replace="gcc"
complement(4820. .5356)
/citation=8
/codon_start=1
/transl_table=11
/product="YFHB-ECOLI protein"
/protein_id="BA10910.1"
/db_xref="GI:987637"
/translation="MRARITGVTFPLSEVFSHEVYMRALTLAKRAMPREYGVAV
LYNNRVIGSGMRPIGRHDPTVAHEIMLRGGGLVMONVRLIDTLVYTLRPGWCA
GANHTSRIGRVFGARDATGAAGSLMDVLRHGMNRVETITGILADECALLSDPF
RMKRQRIKQKAQSTSD"
complement(5381. .5995)
/citation=8
/codon_start=1
/transl_table=11
/product="YFHB-ECOLI protein"
/protein_id="BA10910.1"
/db_xref="GI:987638"
/translation="WVFPLDGTILHQDQFSGFLRYLRROPALNALLVPLPIIAIA
LIIKGRARPMPSLIMGCTFGSSEARLQTLQADFYVRPRDVTVAPLVOERLTYLL
SSADLIWLTGSRQPLVEAVYPTPWLPRVNLJASQIDRGYGGWVLTWRCIGHKXAQ
LERKIGTPRLVSGYSDSNQNDPLIFCQRRRVTPRGLQLE"
6153. .7073
/citation=8
/codon_start=1
/transl_table=11
/product="YFHH-ECOLI protein similarity"
/protein_id="BA10911.1"
/db_xref="GI:987639"
/translation="MEYKHSRVSHIACRRHYFIARKDNGLLRIRORYQGLAQSDK
LADYLLQDPTVAHLSQQLANAGVSSVFAQKLGKGPALAKLISEMLASQP
ESSVSVINQIRGDDPLRVEKLIKENRAAMVATLNNSSEKLCHECYMLISARLI
LTGIGSGVAGNCFAMKMKIGBNAAVADHALLATVQASSDDLLIISYGVARE
LNIADBNMRVGSKVLTITGFTPNALQQAASHCLVYTAEQATNSISACHAQSLT
DLFLALIQDDELAPERIRHSEALVKLV"
complement(9209. .9937)
/citation=7
/codon_start=1
/transl_table=11
/product="reco protein"
/protein_id="BA10912.1"
/db_xref="GI:987640"
/translation="MEGQRAFYLSRPSWSETSLMDVFTBESGRVRLVAKGARSRS
TKGALQPTPLLRFGRGVYTLASAVALPLSGITISGLYINELLSRVLEY
ETRFSLRFPDYLIHCTIOSLAGVSTPEPALRFEALIGHLYGVNFTHCAGSGEPVD
TMYRYRBEKGFILASVINDKFTTGQQLALNAREPDDTLRAARGFTRMALKPVLG
GKDKSRELRQMPKRVITKTYE"
complement(9949. .10854)
gene

Qy	301	GGCAGTATTGACCCGGCAGCGGTACACCTTTTGAGTTGGCGGGCGGCATATATGTT	360
Db	13636	GGCAGTATTGACCCGGCAGCGGTACACCTTTTGAGTTGGCGGGCGGCATATATGTT	13695
Qy	361	CTGGCTGGCGTGGAAAAATCGCACACAGCCCAACAAGAAAGACGACTTCAGGCAAAACC	420
Db	13696	CTGGCTGGCGTGGAAAAATCGCACACAGCCCAACAAGAAAGACGACTTCAGGCAAAACC	13755
Qy	421	AATACGCTTTTGGCCGACGCTTTGCTTGGCACTTTGTCAAGTCAATCATTTTGTACGG	480
Db	13756	AATACGCTTTTGGCCGACGCTTTGCTTGGCACTTTGTCAAGTCAATCATTTTGTACGG	13815
Qy	481	TGTTACGGCACTGTCGACGTTTGTCTGCGCAAAACAGAGCGTTAAGCTGGTAGTTGG	540
Db	13816	TGTTACGGCACTGTCGACGTTTGTCTGCGCAAAACAGAGCGTTAAGCTGGTAGTTGG	13875
Qy	541	CGTCAGCGTTTGTGCTGCGATGATTTGGACGTTTGGCAATGTGTGCTGGCGCTGGCGGG	600
Db	13876	CGTCAGCGTTTGTGCTGCGATGATTTGGACGTTTGGCAATGTGTGCTGGCGCTGGCGGG	13935
Qy	601	GCATCTGTTTCAGCGATTTGTTGCCGATATGTCGCCAGTAAATATCGTGTGCTGCCCT	660
Db	13936	GCATCTGTTTCAGCGATTTGTTGCCGATATGTCGCCAGTAAATATCGTGTGCTGCCCT	13995
Qy	661	GTTGCTGATCTATTGGCGCGGATGACATTTTCTATTAAAGAAAAAAGCGAAGAGTCCGC	720
Db	13996	GTTGCTGATCTATTGGCGCGGATGACATTTTCTATTAAAGAAAAAAGCGAAGAGTCCGC	14055
Qy	721	CTTCTTCGCTTAGTAACTTGCTACTTAAG	750
Db	14056	CCTCTCGCTTAGTAACTTGCTACTTAAG	14085
RESULT 6			
AE015279			
LOCUS	Shigella flexneri 2a str. 301	12322 bp	DNA linear BCT 18-OCT-2002
DEFINITION	Shigella flexneri 2a str. 301 section 242 of 412 of the complete genome.		
ACCESSION	AE015279	AE005674	
VERSION	AE015279.1	GI:24053018	
KEYWORDS	Shigella flexneri 2a str. 301		
SOURCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.		
ORGANISM			
REFERENCE			
AUTHORS	1 (bases 1 to 12322)		
	Jin, Q., Yuan, Z. H., Xu, J. G., Wang, Y., Shen, Y., Lu, W. C., Wang, J. H., Liu, H., Yang, J., Yang, F., Qu, D., Zhang, X. B., Zhang, J. Y., Yang, G. W., Wu, H. T., Dong, J., Sun, L. L., Xue, Y., Zhao, A. L., Gao, Y. S., Zhu, J. P., Kan, B., Chen, S. X., Yao, Z. J., He, B. K., Chen, R. S., Ma, D. L., Qiang, B. Q., Wen, Y. M., Hou, Y. D. and Yu, J.		
TITLE	Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157		
JOURNAL	Nucleic Acids Res. 30 (20), 4432-4441 (2002)		
PUBMED	12384590		
REPLACE			
AUTHORS			
TITLE	2 (bases 1 to 12322)		
JOURNAL	Jin, Q., Shen, Y., Wang, J. H., Liu, H., Yang, J., Yang, F., Zhang, X. B., Zhang, J. Y., Yang, G. W., Wu, H. T., Dong, J., Sun, L. L., Xue, Y., Zhao, A. L., Gao, Y. S., Zhu, J. P., Lu, W. C., Qiang, B. Q., Wen, Y. M. and Hou, Y. D.		
PUBMED	12384590		
REPLACE			
AUTHORS			
TITLE	Direct Submission		
JOURNAL	Submitted (08-MAY-2003) Microbial Genome Center of Chinese Ministry of Public Health, 100 Yingxin Jie, Xuanwu Qu, Beijing 100052, P.R. China		
FEATURES			
source	Location/Qualifiers		
	1..12322		
	/organism="Shigella flexneri 2a str. 301"		
	/mol_type="genomic DNA"		
	/strain="301"		
	/serotype="2a"		
	/db_xref="taxon:198214"		
	218..1840		
	gene		

CDS			
	/gene="nadB"		
	/locus_tag="SF2636"		
	/pseudo		
	218..1840		
	/gene="nadB"		
	/locus_tag="SF2636"		
	/note="Similar to Escherichia coli K12 quinolinate synthetase, B protein gi: 1788928 (541 aa). BLAST with identity of 97% in 541 aa. This CDS contains an in-frame stop codon. The sequence has been checked and is believed to be correct."		
gene	/pseudo		
	/codon_start=1		
	/transl_table=11		
	complement(1825..2682)		
CDS	/gene="yfiC"		
	/locus_tag="SF2637"		
	complement(1825..2682)		
	/gene="yfiC"		
	/locus_tag="SF2637"		
	/note="Residues 1 to 285 of 285 are 98 pct identical to residues 1 to 285 of a 285 aa protein from Escherichia coli K12 ref: NP_417070.1"		
	/codon_start=1		
	/transl_table=11		
	/product="putative enzyme"		
	/protein_id="AA044133.1"		
	/db_xref="GI:24053019"		
	/translation="MMGADYRRYGCNLIHPWISLIGYSHVLSPEFLIOGTFINQS TSVLRNGPFPFOFVADRCAMKGTGDIILGAMAPVAGVRCIDIGAGSLMLM AGRSDSVIIDAVEISEAAAOENINOSPWAERINVTADIQWITQOTVRFDLI ISNPPYQOQVECAETPORBOARYTTLDHPSLTLCAACITBEGFECVLPEDINGFT ELASMGHLRLRTDVAENEARLPHRVLLAFSPQAGECSDBLVRGPQONSAYATA LRGAFYLPW"		
	2694..4028		
	/gene="srnB"		
	/locus_tag="SF2638"		
	2694..4028		
CDS	/gene="srnB"		
	/locus_tag="SF2638"		
	/note="Residues 1 to 444 of 444 are 100 pct identical to residues 1 to 444 of a 444 aa protein from Escherichia coli O157:H7 ref: NP_311469.1"		
	/codon_start=1		
	/transl_table=11		
	/product="ATP-dependent RNA helicase"		
	/protein_id="AA044134.1"		
	/db_xref="GI:24053020"		
	/translation="MTVTFSLELDESLEALQDKGFTRPATIAQAALPPLADGADV LGSAPPTGKTAAYLLPALQHLIDFPRKSGPPRLILPTRELAQAQVADHARELAKH THLDIATITGGVAVYMHAEVSESNODIVATTGRLQYIKENFDCAVEETILDEAD RMDMGFPADIDIEHIAGETRRKOTLIFSATIGSDAIDQRAERLDEPVSANPSIRE RKKITHMYTRADLIEKTLALVHLILQKPATRSIVYRKGRVHELAAMLREAGINRC YLEGENVOGRNEAIRLTGRVNVAVTDVAARGIDIPDVSHVENFMDPRSGDTLH RIQRTARAGKGTALISLVEAHDLILGKGRVIEPIKARVIDELRPTKRAPEKQYG KPSKVTIARAEKKEKKEKPRVKGRHBDYNIKGRRRSGTGVPQTTEE"		
	complement(4063..4989)		
gene	/gene="yfiB"		
	/locus_tag="SF2639"		
	complement(4063..4989)		
CDS	/gene="yfiB"		
	/locus_tag="SF2639"		
	/note="Residues 1 to 306 of 308 are 96 pct identical to residues 1 to 306 of a 308 aa protein from Escherichia coli O157:H7 EDJ933 ref: NP_289135.1"		
	/codon_start=1		
	/transl_table=11		
	/product="putative transcriptional regulator LYSR-type"		
	/protein_id="AA044135.1"		
	/db_xref="GI:24053021"		
	/translation="WFLYSKPDLPDRFGVMDLRRTTLTKTVVEGSEFLASQYLCTG STVTFHIQQLQEGFSVQLFEKIGRWCLTRREGKLLPHIYELTRVMDTLREAKKSD PDELRVSGETLISYRMPQVLFRRFQAPAKVRLSLQALNCYVIRDALINDEADVGF		

RESULT 7
AE016987
LOCUS
DEFINITION
Shigella flexneri 2a str. 2457T section 10 of 16 of the complete genome.
ACCESSION
AE016987 AE014073
VERSION
AE016987.1 GI:30042162
KEYWORDS
Shigella flexneri 2a str. 2457T
Shigella flexneri 2a str. 2457T
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.
REFERENCE
1 (bases 1 to 290380)
Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W., Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C. and Blattner, F.R.
Complete Genome Sequence and Comparative Genomics of Shigella flexneri Serotype 2a Strain 2457T
Infect. Immun. 71 (5), 2775-2786 (2003)
JOURNAL
12704152
2 (bases 1 to 290380)
Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W., Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C. and Blattner, F.R.
Direct Submission
Submitted (13-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
TITLE
JOURNAL
source
Location/Qualifiers
1. 290380
/organism="Shigella flexneri 2a str. 2457T"
/mol_type="genomic DNA"
/strain="2457T"
/serotype="2a"
/db_xref="taxon:198215"
complement (82. 1209)
/gene="gcpB"
/locus_tag="S2733"
/note="Residues 1 to 372 of 375 are 96.23 pct identical to residues 1 to 372 of 372 from Escherichia coli K-12 : B2515"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAP17889.1"
/db_xref="GI:30042163"
/translation="MENOAPIQRRKSTRIVYGNVPIDGAPIAVQSMTRTTDVEATVNCIKALERVAGDIVRVSVPTMDAAEFKIKOVNVNPLVADHFDYRIALKAVEYVDLRIKINPENGIERIRIMVVDCAKDKNIPIRIGVNASLEKIDQEKYGEPTQALLSSAMRVVDHLDFLNFOFKVSKASDVPLAVESYRIILAKOIDPLHGITTEAGASGAVKSAIVGILLSEGIQDITRVSLADPVEIKVGGDIILSKLRISRGINPLIACPTCSGQEFVDTGYNALRQRLIEDITTPMDYSIIICGVNNGGELVSTLRTGTGKKSGLYEDVRKDRIDNDMDIDQELARIRAVASQDDEARIDVQVQVKITIT"

gene
complement (2534. 3688)
/gene="yfgB"
/locus_tag="S2735"
complement (2534. 3688)
/gene="yfgB"
/locus_tag="S2735"
/note="Residues 1 to 384 of 384 are 99.47 pct identical to residues 1 to 384 of 384 from Escherichia coli K-12 : B2517"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAP17890.1"
/db_xref="GI:30042165"
/translation="MSEQLVTPENVTYDKKINILDLNRQWRFFPKDGEKTPRADOVKMWMYHCCDNFDEDTIDIKVRLKLVAEITRAPEVEGRSSDGTIKAIANGDQREVTYIPEDDRALTLCSVSQVGCALCKFCSTAQQGFNRNLSVSIIGQWRAKIVGAAKVTGQRPITVVMVMGGEPLNNDNVNPAEIMLDGFGELSRKRYTLSTGVVPA LRLGMDIVLALISLHAPNDIEIRIYINKYNIETFLAVRRLKSNAMQCRVTIRVYMLDHVNDGTEHAHOLAEILKDPCKINLI PNVPFDAYGRSSNSRIDRFKVLMSYGPITVYRKTRGPDIDACQGLAGVDYDRKRLRLRQGEALADIRAV"
complement (3838. 4269)
/gene="ndk"
/locus_tag="S2736"
complement (3838. 4269)
/gene="ndk"
/locus_tag="S2736"
/function="enzyme, Purine ribonucleotide biosynthesis"
/note="Residues 1 to 143 of 143 are 100.00 pct identical to residues 1 to 143 of 143 from Escherichia coli K-12 : B2518"
/codon_start=1
/transl_table=1
/product="nucleoside diphosphate kinase"
/protein_id="AAP17891.1"
/db_xref="GI:30042166"
/translation="MAIERTFSLIKRNAVAKKVIIGNIFRPAAGFKIVGTQMLHLYV EQRGPFARHDKGPFEDGI VEFMTSGPIVSLVEGNVAQRHDLGATNPANALAGT LRDYVDSITENGTHSGDSVESAAKRIATFRBGEVCPRTR"
complement (4418. 6517)
/gene="ppc"
/locus_tag="S2737"
/note="truncated"
/pseudo
6514. 7961
/note="insertion element"
/insertion_seq="ISSF14"
complement (6534. 7730)
/locus_tag="S2738"
complement (6534. 7730)
/locus_tag="S2738"
/function="IS, phage, Tn; Transposon-related functions"
/note="Residues 1 to 398 of 398 are 90.95 pct identical to residues 1 to 398 of 398 from GenPept : >gb|AAK18438.1|AF348706.127 (AF348706) putative transposase for IS110 [Shigella flexneri]"
/codon_start=1
/transl_table=1
/product="ISSF14 orf"
/protein_id="AAP17892.1"
/db_xref="GI:30042167"
/translation="MTSSSDYBSVOVFIGVDGKDTTHHVAIINRSGRLFDKALPNDS NKIRSLISDKQGGQILVVDOPATIGAPVAVARSEGVLVGLPGLAMRRIADLHAG EAKTVRDAAITIAEARKTLPHALRTIKLDEQIAELSMICGPPDDLAQTQASNRIR GLTLOIHAPERVLGRLEHPAVLDLQRYPSSEKIASIGKKLAQALQVRLRGR LAAIDIAQALAEQTVVVPVTMAAVVPLRLALQILTRKORDEVALVEORVLAHPLYP

gene
complement (1236. 2249)
/locus_tag="S2734"
complement (1236. 2249)
/gene="yfgA"
/locus_tag="S2734"
/function="putative membrane; Not classified"
/note="Residues 1 to 337 of 337 are 80.71 pct identical to residues 1 to 337 of 337 from Escherichia coli K-12 : B2516"
/codon_start=1
/transl_table=1
/product="putative membrane protein"
/protein_id="AAP17889.1"
/db_xref="GI:30042164"

repeat_region
gene
complement (1236. 2249)
/locus_tag="S2734"
complement (1236. 2249)
/gene="yfgA"
/locus_tag="S2734"
/function="putative membrane; Not classified"
/note="Residues 1 to 337 of 337 are 80.71 pct identical to residues 1 to 337 of 337 from Escherichia coli K-12 : B2516"
/codon_start=1
/transl_table=1
/product="putative membrane protein"
/protein_id="AAP17889.1"
/db_xref="GI:30042164"

OY	181	TATTCCTGGCCCTTAGCTCTGCTACAGTGCATGGAATTTCGTCAAAGTACCCGCGTCTGGC	240
Db	82342	TATTCCTGGCCCTTAGCTCTGCTACAGTGCATGGAATTTCGTCAAAGTACCCGCGTCTGGC	8240
OY	241	AGGATGATGTCGTGGGAATTTTGTGATTTGATGATTAAGTAACTGATGCGGGGCAATTCATTTTCACT	300
Db	82402	AGGATGATGTCGTGGGAATTTTGTGATTTGATGATTAAGTAACTGATGCGGGGCAATTCATTTTCACT	8246
OY	301	GGCAGTATGTAACCCCGCAGCGGTACAACCTTTTGATTTGGGCGGGGCGCATATATTTGT	360
Db	82462	GGCAGTATGTAACCCCGCAGCGGTACAACCTTTTGATTTGGGCGGGGCGCATATATTTGT	8252
OY	361	CTGGCTGGCGCTGGAAAAATCGCCACAGCCCCAACAAAGAAAAGACGACTTCAGCAAAACC	420
Db	82522	CTGGCTGGCGCTGGAAAAATCGCCACAGCCCCAACAAAGAAAAGACGACTTCAGCAAAACC	8258
OY	421	AATACGCTTTTGGGCCGACCTTTGCTTTGCACTTTGCAATGCTCAAAATTCATTTTGTACGG	480
Db	82582	AATACGCTTTTGGGCCGACCTTTGCTTTGCACTTTGCAATGCTCAAAATTCATTTTGTACGG	8264
OY	481	TGTTACGCGCACATGTCGACAGTTTGTTCCTGCGCAAAACACAGGCGTTAAGTGGATGTGG	540
Db	82642	TGTTACGCGCACATGTCGACAGTTTGTTCCTGCGCAAAACACAGGCGTTAAGTGGATGTGG	8270
OY	541	CGTGAGCGCTTTTGTCTGGCGATGATTTGGGAAGCTTTGGCAATGTGTCTGTGGCGCTGCGG	600
Db	82702	CGTGAGCGCTTTTGTCTGGCGATGATTTGGGAAGCTTTGGCAATGTGTCTGTGGCGCTGCGG	8276
OY	601	GCATCTGTTTCAGCGAATTTTGTGGCCACGATATGTGTCCGCACTTAATATCGTCTGTCCCT	660
Db	82762	GCATCTGTTTCAGCGAATTTTGTGGCCACGATATGTGTCCGCACTTAATATCGTCTGTCCCT	8282
OY	661	GTTCCTGATCTATTTGAGCGCGGTACGCACTTTCTATTAACGAAAAAAGCGAGAAGTCTGC	720
Db	82822	GTTCCTGATCTATTTGAGCGCGGTACGCACTTTCTATTAACGAAAAAAGCGAGAAGTCTGC	8288
OY	721	CTTCTTCGCGCTAGTAATACTTGCACTTAAG	750
Db	82882	CTTCTTCGCGCTAGTAATACTTGCACTTAAG	82911
RESULT 8			
LOCUS	AE016764	300099 bp DNA linear BCT 24-DEC-2002	
DEFINITION	Escherichia coli CFT073 section 10 of 18 of the complete genome.		
ACCESSION	AE016764 AE014075		
VERSION	AE016764.1 GI:26109116		
KEYWORDS	.		
SOURCE	Escherichia coli CFT073		
ORGANISM	Escherichia coli CFT073 Bacteria; Proteobacteriia; Gammaproteobacteriia; Enterobacteriales; Enterobacteriaceae; Escherichia. 1 (bases 1 to 300099)		
REFERENCE	Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P., Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,U., Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C., Perna,N.T., Mobley,H.L.T., Domnenberg,M.S. and Blattner,F.R. Extensive Kioske Structure Revealed by the Complete Genome Sequence of Uropathogenic Escherichia coli Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)		
JOURNAL	PUBLISHED	12471157	
REFERENCE	2 (bases 1 to 300099)		
AUTHORS	Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P., Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,U., Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C., Perna,N.T., Mobley,H.L.T., Domnenberg,M.S. and Blattner,F.R. Direct Submission		
TITLE	Submitted (20-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA		
JOURNAL	Location/Qualifiers		
FEATURES	1. .300099		
SOURCE	/organism='Escherichia coli CFT073'		

```
gene
/mol_type="genomic DNA"
/strain="CFR073"
/db_xref="taxon:199310"
complement(163. .1383)
/gene="fabB"
/locus_tag="c2869"
complement(163. .1383)
/gene="fabB"
/locus_tag="c2869"
/EC_number="2.3.1.41"
/function="enzyme; Fatty acid and phosphatidic acid biosynthesis"
/note="Escherichia coli K-12 ortholog: b2323; Escherichia coli O157:H7 ortholog: z3586"
/codon_start=1
/transl_table=11
/product="3-oxoacyl-[acyl-carrier-protein] synthase I"
/protein_id="AA081319.1"
/db_xref="GI:26109117"
/translation="MKRAVITGLIVSSIGNNOEVLASLREGSGITFSGELKDSGM RSHVGNVXKIDTGLIDRKVYRPMSSDSITAPLMEQATADAGLSPAYONNPRVGLI AGSGGSPRFQVFGADMRGPRGKAVGPVVTIKAMSGVACIATPFKIHGNYIS SACATSAHICIGNAVEQIQKQDIFVAGGEBLCWEMACFPDAGALSTKYNTPESA SRTYDARDGFVLAGGGVWVELEHALRGAHVAIVGYATSDGADVAPSGG AVRCMKAMHGVDTPIDYNSHGTFSTGVKELAIREVFGDKSPAIISATKMTGHS LGAAGVQEAIVSLMLEHGFAPSIINIEIDEQAGLNIVTETDRELTVMSNSFGF GGNATLVMKRLMD"
1482. .3548
/gene="yfcM"
/locus_tag="c2870"
1482. .3548
/gene="yfcM"
/locus_tag="c2870"
/function="putative"
/note="Escherichia coli K-12 ortholog: b2324; Escherichia coli O157:H7 ortholog: z3587"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yfcM"
/protein_id="AA081320.1"
/db_xref="GI:26109118"
/translation="MRSILTHAKIEPPVRVRYVYKHYSIQPANLEFNAEGTPVSRDF DDVYFSNDNGLEETRYVFLGNNLEARPEHPPLPVVASGFGTGINPLTQAPQO FREAHPOAOLRHLHIFSEKPEPLRADLALAHOMPELAPABOLQOMPLPQGCGR LLDGEGHTLIDMGGDINELTSDLSLNKVAMFLDGFAPAKNPMPTOMONLENMAA RLARPGSTLATFTSAGFVRGGLDQAGFTMQKRGFKRKEMLCGVMQTIPRESTW FNRWGSNKQKEAIIIGGIASALSLALLRRGMQVTLVCADEAPALGASGRQGLVPL LSKHDEALNRFNSAFTFARFYDLPVKCDHMCVTOLGWDEKSHKLAOMLSMDL PAELAVAVEANAQVQITGVATNCSGITVPOGWCSPALTRNVLELAQQOGLQIHYH QLODLSKDGWMLNFAGDHATHSVVVLANGHOISGFSOTSLPVYSVAGVSHIPT TPELAKIKQVLCYNGVLTBPONRANOHHCIGASVYRGSEETA SPEDQONRORLIDF PVOVMKEADVSGKEARCGVRCATRDHLPMVGNVPDYDAILVERASIAEKDEVNSAP VYDLEFMPALGSRGLCSAPLCAEIILAOWSEEPDPMDASTLALNPNRLMVRKLLG KAVVAG"
```

```
gene
/locus_tag="c2872"
/function="putative"
/note="Escherichia coli K-12 ortholog: b2326; Escherichia coli O157:H7 ortholog: z3589"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yfcM"
/protein_id="AA081322.1"
/db_xref="GI:26109120"
/translation="MNSTHYEQLIEIFNSCPADEPNTLIKGDDEPIYLPADAEPY NRIVPAHGTYASAIHEISHTICAGKARRLVDGYYCCPDGAPQOSQPEDVEVQO ALDWLFCVAGVYFNVSCNLSGDFEPDVFQRVHAQVMDYLAINGIIPERPARFIKA LQNYVTPBELTAEQFPWPEALN"
complement(4529. .5338)
/gene="yfcA"
/locus_tag="c2873"
complement(4529. .5338)
/gene="yfcA"
/locus_tag="c2873"
/function="putative structure"
/note="Escherichia coli K-12 ortholog: b2327; Escherichia coli O157:H7 ortholog: z3590"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yfcA"
/protein_id="AA081323.1"
/db_xref="GI:26109121"
/translation="METPNSLFWVSPLLIGVLFVVMAGLFTDSINGGGILITPILM AAGMSPANALATNKLQACGSGISATVITRRKVSLSQDKLNTAMTFVSGSGALLVQ YVQADVLRQILPLIVICIGLYFLMKLEEDRQRMVGLPFLALAGGCVGFYDGFPG PAAGSFYALAFVTLGCFNLAKTAHAKMLNATSNIGLLIFLIGGVIVATGVVMLVG OFLGARWGSRLVYSKGOCLRPWIVIVASVMSAKLLYDSHQEILHMLGN"
complement(5338. .6162)
/gene="mepA"
/locus_tag="c2874"
complement(5338. .6162)
/gene="mepA"
/locus_tag="c2874"
/EC_number="3.4.99.-"
/function="enzyme; Murein sacculus, peptidoglycan"
/note="Escherichia coli K-12 ortholog: b2328; Escherichia coli O157:H7 ortholog: z3591"
/codon_start=1
/transl_table=11
/product="penicillin-insensitive murein endopeptidase precursor"
/protein_id="AA081324.1"
/db_xref="GI:26109122"
/translation="MNKTAIALALALASSASLAAATPMOKITQVPGSAOSIGSFNSGC IVADTLPIOSEHYOVMRTDORRYFGHPDLVWFLOLSRQVSLMGKTVLIGMGMPA GGRFNGHASHQGLDVIDFLQPKTRWISAOQLLRQALDIVSRDGGKHAIVALKRPA FSLIKLAQDKDVTIRLIVNPATIQOCLDGTDRDMVRKVRPWFQRRAHMVAIRCPA DLSLECDQPLPPPGDGCALQSWFEPKPTTPEKTPPPLPSCQALLDDEHVI"
complement(6166. .7254)
/gene="aroC"
/locus_tag="c2875"
complement(6166. .7254)
/gene="aroC"
/locus_tag="c2875"
/EC_number="4.2.3.5"
/function="enzyme; Amino acid biosynthesis: Chorismate"
/note="Escherichia coli K-12 ortholog: b2329; Escherichia coli O157:H7 ortholog: z3592"
/codon_start=1
/transl_table=11
/product="Chorismate synthase"
/protein_id="AA081325.1"
/db_xref="GI:26109123"
/translation="MMAGNTIGQLFRVTTFGSGHLAGCTIVGVNPGPRTPLTADLOH DLDRRRGTSRYTQRRPQVKILSGVEGVTTGTSIGLLINTQORSODYSALIKVY FRPGADYTEQYKQGLDYRGGRSSARETAMVVAAGALAKKYLAEKFLIEIGCLTQ MGDIPLIEIKDWSQVEQNPFPCPDPRKIDALDELIRALAKKEGDSIGAKVTYVAGSVAG
```

CDS

gene

```
complement(3981. .4529)
/locus_tag="c2872"
complement(3981. .4529)
```

CDS

gene

```
complement(3669. .3947)
/locus_tag="c2871"
complement(3669. .3947)
```

CDS

gene

LGEVPEPDLRLADIAHMLMSINAVGEIGDPVVALRGSQNEDEITKQFQSNHAG
 ILGSSGQOIIAHMALKPTSSITVPERITNRFGEVEEMITKGRHDPVCGIRAVPIAE
 AMALVIMDLRLORANADVKDIPRW"
 complement (7286. .8551)
 /gene="yfcB"
 /locus_tag="c2876"
 complement (7286. .8551)
 /gene="yfcB"
 /locus_tag="c2876"
 /EC_number="2.1.1.72"
 /function="putative"
 /note="Escherichia coli K-12 ortholog: b2330; Escherichia
 coli O157:H7 ortholog: z3593"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical adenine-specific methylase yfcB"
 /protein_id="AN81326.1"
 /db_xref="GI:26109124"
 /translation="MAGIGLPLNPLWHPADPFTLRSDVNRVILNLASPRHOLAKOR
 LVLLAKCCFLFHLSPSISNPEPTVPTSTDRKRIEPHSPRVNADLSSTQTSRN
 YCHALREYVDKIFVDEAVNELIOTIDMLFMSVRSRANITWGHGTNDPDAVOLV
 LPSLYLPDIPEDMRTARLTSSSEKRIIVERVIRVNRKIPVALITNRAMPGHEFYD
 ERVLVRSPIGELINKFAGLISKQPHQIDMCTGSCGLAIACAVAFPEAEVDAVIS
 PDALVAEONIEBHGILHNVIPIRSDIFRDLPRVQYDLITNPPEYDAEDMSDLPRY
 RHEPRLASGTGDKLTRRIIGNADYLDADGVLCVENSVMHMEQYDVPEFTWL

Query Match 98.54; Score 738.8; DB 1; Length 300099;
 Best Local Similarity 99.18; Pred. No. 2.2e-220;
 Matches 743; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

1 GATCATTAACCCCAACCTATCGAATATCGAATATGAAATATTAACATTCATTTT 60
 255985 GATCATTAACCCCAACCTATCGAATATCGAATATGAAATATTAACATTCATTTT 256044
 61 TTAATGTTCCGTGTCGGTACTGCTCAACAAAACAGAGAGATTAACAGTGACCCGAC 120
 256045 TTAATGTTCCGTGTCGGTACTGCTCAACAAAACAGAGAGATTAACAGTGACCCGAC 256104
 121 CCTTTAAGTCCTTTTGAACCTTACACCCCTGATTACCGCTATGACCGCAGACCGAACA 180
 256105 CCTTTAAGTCCTTTTGAACCTTACACCCCTGATTACCGCTATGACCGCAGACCGAACA 256164
 181 TATTCGACCCCTAGCTCTGCTACGTGAGTGAATTCGCAAGTACCCGCGTGGC 240
 256165 TATTCGACCCCTAGCTCTGCTACGTGAGTGAATTCGCAAGTACCCGCGTGGC 256224
 241 AGGATGAGTCTGGGATTTTGAATGTGATGATTAAGTGTGCGGCGCATTCATTTCACT 300
 256225 AGGATGAGTCTGGGATTTTGAATGTGATGATTAAGTGTGCGGCGCATTCATTTCACT 256284
 301 GGCAGTATGATGACCCGCGACCGGTACACCTTTGAGTGGCGCGGCGGCATATATTTGT 360
 256285 GGCAGTATGATGACCCGCGACCGGTACACCTTTGAGTGGCGCGGCGGCATATATTTGT 256344
 361 CTGGCTGGCGTGGAAAATGCGACCAAGCCCAACAAAGGAAGAGCGGACTTCAGGCAAAAC 420
 256345 CTGGCTGGCGTGGAAAATGCGACCAAGCCCAACAAAGGAAGAGCGGACTTCAGGCAAAAC 256404
 421 AATCAGCTTTTGGCGCAGCTTTGCTTGAAGTGTGAAAGTCAAAATCAATTTTGTACGG 480
 256405 AATCAGCTTTTGGCGCAGCTTTGCTTGAAGTGTGAAAGTCAAAATCAATTTTGTACGG 256464
 481 TGTTAAGGCACTGTCCAGCTTTGTTCTGCGGCAACACAGGCGTTAAGCTGGGTAAGTGG 540
 256465 TGTTAAGGCACTGTCCAGCTTTGTTCTGCGGCAACACAGGCGTTAAGCTGGGTAAGTGG 256524
 541 CGTCAGGCTTTGTCGCGATGATTTGGGAGCTTTGGCAAGTGTGCGTGGCGCGTGGGCGG 600
 256525 CGTCAGGCTTTGTCGCGATGATTTGGGAGCTTTGGCAAGTGTGCGTGGCGCGTGGGCGG 256584
 601 GCATCTGTTTCAAGCGATTTTTCGCGAGTATGTCGCGCAGTTAAATATCTGCTTGGCCT 660
 256585 GCATCTGTTTCAAGCGATTTTTCGCGAGTATGTCGCGCAGTTAAATATCTGCTTGGCCT 256644

QY 661 GTTGCTGTCTATTTGGCGCGGTACGACATTTTCTATTACGAAAAAGCGGAGAGGTGCG 720
 Db 256645 GTTGCTGTCTATTTGGCGCGGTACGACATTTTCTATTACGAAAAAGCGGAGAGGTGCG 256704

QY 721 CCTCTTCGCGCTTAGTACTGACTTAAAG 750
 Db 256705 CCTCTTCGCGCTTAGTACTGACTTAAAG 256734

RESULT 9
 AE005488
 LOCUS
 DEFINITION
 of 280.
 ACCESSION
 AE005488 AE005174
 VERSION
 AE005488.1 GI:12516994
 SOURCE
 ORGANISM
 Escherichia coli O157:H7 EDL933
 Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 107
 of 280.
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 1 (bases 1 to 12337)
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grodbeck,E.J., Davis,N.W., Lim,A., Dimalanta,B., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.
 Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
 Nature 409 (6819), 529-533 (2001)

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 11206551
 2 (bases 1 to 12337)
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grodbeck,E.J., Davis,N.W., Lim,A., Dimalanta,B., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.
 Direct Submision
 Submitted (22-OCT-2000) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES
 source
 1. .12337
 /organism="Escherichia coli O157:H7 EDL933"
 /mol_type="genomic DNA"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155864"
 /note="enterohaemorrhagic"
 235. .1857
 /gene="nadB"
 /note="synonym: z3856"
 235. .1857
 /gene="nadB"
 /function="enzyme, Biosynthesis of cofactors, carriers:
 pyridine nucleotide"
 /note="Residues 1 to 540 of 540 are 99.07 pct identical to
 residues 1 to 540 of 540 from Escherichia coli K-12 Strain
 MG1655: B2574"
 /codon_start=1
 /transl_table=11
 /product="quinolinate synthetase, B protein"
 /protein_id="AAG57690.1"
 /db_xref="GI:12516995"
 /translation="NWTLREHSCDVLIIGSGAAGLIALRLADQHOVTLISKPVTEG
 STFYAGGIAAVFDETDSDISVETDLIAGATICRBAVEFPYASNAKRSVOWLIDGVS
 LPTHTIQNGEESYHITREGGSHRRILIAADATGEBVTLVASKALNPRIKLVRS
 NAVDLIISDKIGLPGTRRVAGAVWNRNKEVETCAKAVIATAGAGAVLYQYTNPD
 ISSGDIAMAKPAGCVANLBNQPHPTLALYPOANFLTETALGEGAVLYRPGDTR
 FMPDFVRGELAPRDIVARAIDHEMKRLCADCPFLDISKPADPFIROHPPMIYEKLLG
 LDIDLQBEVPIVPAHYTCGGVMVDHQRTEVBEGLYALGEVSYGLHGANMANSLS

LECLVYGSNAADITRRMPYADHISTLPWDESRVENPDERVIIQHNMHELRLFMWDY
VGIVRTTKLERALRITMLQOEIDEXAHFRVSNMLLELRLNVQVAELIVRCAMARK
ESRGHFTLDYPELLTHSGPSLSPGNHYINR"
/complement (1842. .2699)
/gene="yfiC"
/note="synonym: Z3857"
/complement (1842. .2699)
/gene="yfiC"
/function="putative enzyme: Not classified"
/note="Residues 1 to 285 of 285 are 95.08 pct identical to
residues 1 to 285 from Escherichia coli K-12 Strain
MG1655: B2575"
/codon_start=1
/transl_table=11
/product="putative enzyme"
/protein_id="AAG57691.1"
/db_xref="GI:12516996"
/translation="MMGADYRRYGCNHLHFWISLKGSHPVLSPPFLQGIEMFSQS
TSVLRNNGFTKQFVADRCAMKVGTDGLGAMAPVAGVKRLDIGSGSLALML
AORTSDSVIIDAVEISEAATQAEENVQSPMERINVHADIQOWTQOTAFDIL
SNPPEYGEVCATPQREOARTTSDHOTLTCACAEICTEBGEFCVLPQOIGNSFT
ELASMGHRLRARTDAENEARLPHVLLAFSQAGCEPFDRLVIRGPDQNYSEAYTA
LTPAFILPM"
/gene="yfiE"
/note="synonym: Z3859"
/complement (2711. .4045)
/gene="yfiE"
/note="synonym: Z3859"
/complement (2711. .4045)
/gene="yfiE"
/function="enzyme: Macromolecule synthesis, modification:
RNA synthesis, modification, DNA transcript'n"
/note="Residues 1 to 444 of 444 are 99.54 pct identical to
residues 1 to 444 of 444 from Escherichia coli K-12 Strain
MG1655: B2576"
/codon_start=1
/transl_table=11
/product="ATP-dependent RNA helicase"
/protein_id="AAG57692.1"
/db_xref="GI:12516997"
/translation="MTVTPESELEDSLESLLEALODKGFTRPAIOAAIPALDGRV
LGSAPCTGCTAAVLLPALQHLDPKPKSGPRILITPTRELMQVADHARELH
THDITATITGVAVMHAEVSEKODIVATTGRLLQITKEENPDCAVNTLLXDEAD
RLMDGPAQDIETIHAGETNRKQTLPSATLEBDALQDFAERLLEDVEVSANPSTRE
RKIHQWYRADLEHKTALIVHLKQPEATRSIVFRRKERVHELAMVLRGGINNC
YLEGMOVGKRNBAIKRLTEGRVAVIATVVAAGIDIPDSHVFNMDPMSGDTYLH
RIGRTAAGKGTASISLEAHDHLLKGVRTIIEPKAVIDELRPTKTAPEKQSG
KPSKVLAKEAKKAKKPKRKYKRRHTYIKGRKRSQGVPRQTTEE"
/complement (4078. .5004)
/gene="yfiE"
/note="synonym: Z3860"
/complement (4078. .5004)
/gene="yfiE"
/function="putative regulator: Not classified"
/note="Residues 1 to 308 of 308 are 96.75 pct identical to
residues 1 to 308 from Escherichia coli K-12 Strain
MG1655: B2577"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator LYSR-type"
/protein_id="AAG57693.1"
/db_xref="GI:12516998"
/translation="MPLASKFDIFDRPGVNDLRRFITLKTVEGSGFLASQKLCTQ
STVTFHIQOLEEFSVOLPEKIGRMCLTREGKLLPHIELTRVMDTLREAAKESD
PDGELRVASGETLISYRPMQVLOFRAPKRVRLSLQNCYVRLADLDEADVAG
RYKQNDALNRRELGEOSLYLVASPOADVTEPRGHNAQCFITNEQCVQIOIFS
TLQORITVENITIELSIESIKRCVAAVNTGSYIPRAVEKELSEGSLIELPGEHSQ
TITAMCAHAGKAVSPAMHTFIQVEDCFLPG"
/gene="yfiF"
/note="synonym: Z3861"
/complement (5062. .5649)
/gene="yfiF"
/note="synonym: Z3861"
/complement (5062. .5649)
/gene="yfiF"
/function="orf: Unknown function"
/note="Residues 1 to 195 of 195 are 97.43 pct identical to

residues 1 to 195 of 195 from Escherichia coli K-12 Strain
MG1655: B2578"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG57694.1"
/db_xref="GI:12516999"
/translation="MTPLLSAPFWYTLITAMTPGPNHIALSSATTGHPHSTRVLA
GMSIGFLYMLCAGISFSLAVIDPAVHLLMAGAAVYVMAMKATGSPTEKDGLOT
KISFWASRLQFQVNVKILVGTALSTFLVLPOTOLSIIVGSVLLAMIGFGNVCW
ALAGHLQRLFRQYGRQNLIVALLILYCAVAFY"
/complement (5705. .6088)
/gene="yfiD"
/note="synonym: Z3862"
/complement (5705. .6088)
/gene="yfiD"
/function="putative enzyme: Energy metabolism, carbon:
anaerobic respiration"
/note="Residues 1 to 127 of 127 are 100.00 pct identical
to residues 1 to 127 of 127 from Escherichia coli K-12
Strain MG1655: B2579"
/codon_start=1
/transl_table=11
/product="putative formate acetyltransferase"
/protein_id="AAG57695.1"
/db_xref="GI:12517000"
/translation="MITGIQITKANDLLNSFWLLDSEKGRKIVAKAGYAEDEV
AVSKGIDIRVPEVPEKPEVEGQHLNVMLRRETFEDAVKHEKPPQLTIRVSG
YAVRNSLTPREQRDVIARTFTBSL"
/gene="yfiG"
/note="synonym: Z3864"
/complement (6393. .7082)
/gene="yfiG"
/function="enzyme: Macromolecule synthesis, modification:
DNA - replication, repair, restr./modific'n"
/note="Residues 1 to 229 of 229 are 99.12 pct identical to
residues 1 to 229 of 229 from Escherichia coli K-12 Strain
MG1655: B2580"
/codon_start=1
/transl_table=11
/product="uracil-DNA-glycosylase"
/protein_id="AAG57696.1"
/db_xref="GI:12517001"
/translation="MANELTWHDVLAEEKQPEFLNTLQTVASEROSGVTIYPQKDV
FNAFRTEIGDVVVLIGDDPYGPGQAGLAFSVAPGATPESLKVLENTIPG
ETRRNNGYLESARQGVLLNTVLTWRAGQASHASIGETPTDKYISLINOBEQV
FLMGSHQKQKATIDKQRHVILKAPHPSPLSAHRFGGCMHFVLANQMLEQGETPI
DMNPVLPAESE"
/complement (7130. .8167)
/gene="yfiF"
/note="synonym: Z3865"
/complement (7130. .8167)
/gene="yfiF"
/function="orf: Unknown function"
/note="Residues 1 to 345 of 345 are 99.42 pct identical to
residues 1 to 345 from Escherichia coli K-12 Strain
MG1655: B2581"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG57697.1"
/db_xref="GI:12517002"
/translation="NMDEKKGSGKGVKVMYVRSDDSDKRTNHPRTGKGGRPGKSGRA
DGRRRARDDKQSPDRDKRVEDPMTVTSRABDDEPEKADHGIGSGFLPIDEVLR
QRAEEDRVGENACQALFQSRPAIVAWFTQSVTRFKBALRMMAANKKAAHVVDVA

Query Match 91.7%; Score 687.6; DB 1; Length 12337;
Best local Similarity 94.8%; Pred. No. 1.5e-204;
Matches 711; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 GATTCATACCCGAACCTATCGAAATATCGAATGTAAACATTGATTTT 60
|||||

Db 4953 GATCATACCCCAAACTATCGAAATATGAAATGTAATTAATTCATTTT 5012

Qy 61 TTAATATGTCGTCGTGGGTAATCTGTAACCAAAACAGAGAGATPAACAGCCGAC 120

Db 5013 TTAATATGTCGTGGGTAATCTGTAACCAAAACAGAGAGATPAACAGCCGAC 5072

Qy 121 CTTTAAAGTCCTTTTGAATTAACACCTGATTACCGCTATGACCGCAGAACAA 180

Db 5073 CTTTAAAGTCCTTTTGAATTAACACCTGATTACCGCAGAACCGAACAA 5132

Qy 181 TATTCGCGCCCTTACGTCCTGCTAGTCGACATGATTCGTAAGTACCGCGGCTGGC 240

Db 5133 TATTCGCGCCCTTACGTCGACATGATTCGTAAGTACCGCGGCTGGC 5192

Qy 241 AGGATGATGTCGCGATTTTGTATGATGATGATGATGATGATGATGATGATGAT 300

Db 5193 AGGATGATGTCGCGATTTTGTATGATGATGATGATGATGATGATGATGATGAT 5252

Qy 301 GGCAGTATGACCCGCGACGGGTAACCTTTGATGAGGCGGCGGCGATATATTTGT 360

Db 5253 GGCAGTATGACCCGCGACGGGTAACCTTTGATGAGGCGGCGGCGATATATTTGT 5312

Qy 361 CTGCTGCGCCCTGGAATATCCGACACCAACCAAAAGAGACGACTTACGCAAAACC 420

Db 5313 CTGCTGCGCCCTGGAATATCCGACACCAACCAAAAGAGACGACTTACGCAAAACC 5372

Qy 421 AATGAGCTTTTGGGCGAGCTTTGCTTGCAGATTTGTGAAGCTCAAAATCATTTTGTACGG 480

Db 5373 AATGAGCTTTTGGGCGAGCTTTGCTTGCAGATTTGTGAAGCTCAAAATCATTTTGTACGG 5432

Qy 481 TGTACGCGACCTGTCAGATTTGTTCTGCGCAAAACAGAGCGTTTAACTGGGTAATTTGG 540

Db 5433 TGTACGCGACCTGTCAGATTTGTTCTGCGCAAAACAGAGCGTTTAACTGGGTAATTTGG 5492

Qy 541 CGTCAGGCTTTTGTGTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

Db 5493 CGTCAGGCTTTTGTGTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 5552

Qy 601 GCATCTGTTTCAAGGATTTTTCGCGCAATGATGATGATGATGATGATGATGATGATGAT 660

Db 5553 GCATCTGTTTCAAGGATTTTTCGCGCAATGATGATGATGATGATGATGATGATGATGAT 5612

Qy 661 GTTGTGTCGTCATTTGCGCGGTAACGATTTTCTATTAACGAAAGAGAGATGTCGC 720

Db 5613 GTTGTGTCGTCATTTGCGCGGTAACGATTTTCTATTAACGAAAGAGAGATGTCGC 5672

Qy 721 CCTTCTCCGCTTGAATTAATGCTTGAATTAAG 750

Db 5673 CCTTCTCCGCTTGAATTAATGCTTGAATTAAG 5702

RESULT 10

AP002562 270365 bp DNA linear BCT 20-MAR-2004

LOCUS Escherichia coli O157:H7 DNA, complete genome, section 13/20.

ACCESSION AP002562 BA000007

VERSION AP002562.1 GI:13362858

SOURCE Escherichia coli O157:H7

ORGANISM Escherichia coli O157:H7

REFERENCE 1

AUTHORS Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatemoto, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C., and Shinagawa, H.

TITLE Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak

JOURNAL Genes Genet. Syst. 74 (5), 227-239 (1999)

MEDLINE 20198780

PUBMED 10734605

REFERENCE

AUTHORS

2

Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M., Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and Hayashi, T.

TITLE

Comparative analysis of the whole set of RNA operons between an enterohemorrhagic Escherichia coli O157:H7 strain and an Escherichia coli K-12 strain MG1655

JOURNAL

SYST. APPL. MICROBIOL. 23 (3), 315-324 (2000)

MEDLINE

PUBMED

20557356

11108008

REFERENCE

AUTHORS

3

Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C., and Shinagawa, H.

TITLE

Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak

JOURNAL

Gene 258 (1-2), 127-139 (2000)

MEDLINE

PUBMED

20564182

11110500

REFERENCE

AUTHORS

4

Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasaki, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.

TITLE

Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12

JOURNAL

DNA Res. 8 (1), 11-22 (2001)

MEDLINE

PUBMED

11258796

21156231

REFERENCE

AUTHORS

5

Hattori, M., Ishii, K. and Shiba, T.

TITLE

Direct Submission

JOURNAL

Submitted (26-JUN-2000) Masahira Hattori, Kitasato Institute for Life Sciences, Kitasato University, Kitasato 1-15-1, Sagamihara, Kanagawa 228-8555, Japan

COMMENT

(E-mail:hattori@genome.le.kitasato-u.ac.jp, URL: http://genome.le.kitasato-u.ac.jp/, Tel: 01-42-778-8194, Fax: 01-42-778-8193)

genome project

This work was done in collaboration with Tetsuya Hayashi, Makoto Ohnishi, Keisuke Nakayama (Miyazaki Medical College), Kozo Makino, Ken Kurokawa, Katsumi Yokoyama, Masashi Tanaka, Takeshi Honda, Tetsu Yasunaga, Hideo Shinagawa (Osaka University), Takahiro Murata (Shinshu University), Chang-Gyun Han, Eichi Ohtsubo, Toru Tobe, Chihito Sasaki (University of Tokyo), Hideto Takami (Japan Marine Science and Technology Center), Naotake Ogasawara (Nara Institute of Science and Technology), Satoru Kuhara (Kyushu University), and supported by the Research for the Future Program of the Japan Society for the Promotion of Science.

Location/Qualifiers

1. 270365

/organism="Escherichia coli O157:H7"

/mol_type="genomic DNA"

/strain="O157:H7"

/sub_strain="RIMD 0509952"

/db_xref="taxon:83334"

/complement(435..1220)

/gene="EC63388"

/complement(435..1220)

/gene="EC63388"

/note="similar to SSB-ECOLI gi|1788871 percent identity 98 in 261 aa (Conserved in E.coli K-12)"

/codon_start=1

/evidence=not experimental

/transl_table=11

/product="enhanced serine sensitivity"

/protein_id="BAB36811.1"

/db_xref="GI:13362859"

/translation="MWLMSETNLELDLEKATPAPRPAFRLLTSTVWPGTAA

OSBAVVEDSALDLQHWKEDGTSVIPFTSLBALQDAVEDQAFVVMPTLFEWTLG

gene
CDS

ETLFLNATLPCKEPMREISLLIGESBNPLSSQETIEGGSRLSLSTSVAPPPQMIDS
LTTRKTKIPKVRFARICSIKENEEAQPMLLGIDADDISEILLIQVAGSVATDPLPDE
PDLCIOVKRGKGSGISHFTTEHIAFPERRRMGFLURDFKONRII"
complement (1354. . 2724)
/gene="EC63389"
complement (1354. . 2724)
/note="similar to PERP_ECOLI gi|1788872 percent identity
99 in 456 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative peptidase"
/protein_id="BAB36812.1"
/db_xref="GI:13362860"
/translation="MPSCNANHLEKSMGCOLAAALLQRITKTMEAKITTLSTOPAD
AMGRKATYSINNDGITLHNGADDLGILQPAARKTGLGHKHQLSGSDMDRCSWA
FMQYKAPEGTRKRWEHPDLDDQROBELNRMTIIDWVRDTINAAEBELGSQLAQRV
DLISNVASDRVVYRTIKGEDBLEQQGYMGLTVGSGSESPVLALDYNPFGDEAPY
ACLVSGKITPDSGGYRISKOTAFMDSKMDOGAAIVYGALAFAATTGINKRWGLFLCC
ANLSGNAGFKGDIITYRNGKRGFEVMNTDAEGRLVLADGLIDASAQRPENIIDAATL
TGAAKTLAGNDVHLAFSPDDLAGERLLASAOENEPFWRLPLAEFRSOLPSNEALNT
NNGSAAYPRGASTANGFLSHFPVENYQGMHLIHDSATYRKAPVQMSAGATGCVRTI
ANULYA"

gene
CDS

complement (2815. . 3015)
/gene="EC63390"
complement (2815. . 3015)
/gene="EC63390"
/note="identical to YRHJ_ECOLI gi|1788873 (Conserved in
E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB36813.1"
/db_xref="GI:13362861"
/translation="MGLKWTDTSREIGEALYDAPDLDPTVRFITDMHQWICLDEFPDD
DPQASMERKILEAILLWLMLDEAE"
complement (3027. . 3362)
/gene="EC63391"
complement (3027. . 3362)
/gene="EC63391"
/note="similar to FER_ECOLI gi|1199991|sp|P25528 percent
identity 100 in 111 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="[2FB-2S] ferredoxin"
/protein_id="BAB36814.1"
/db_xref="GI:13362862"
/translation="MPKIYLPHODLCPDGAIVLEANSGETILDALRNGLIEHACEKE
SCACTCHCTIVAREGFDLSPESECEGDMLDKRMGLPEBSRLSCQARVTDEDLVVEIRP
YTINHAREH"

gene
CDS

complement (3364. . 5214)
/gene="EC63392"
complement (3364. . 5214)
/gene="EC63392"
/note="similar to HSCA_ECOLI gi|1788875 percent identity
100 in 616 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="molecular chaperone"
/protein_id="BAB36815.1"
/db_xref="GI:13362863"
/translation="MALLIQISBPGLSAAPHORRLAAGIDLGTNSLVATVRSQAETTL
ADHEGSLPLSVHYEQOQHGVGYARTMAALDNTANTISSVRKLMBGSLADIQORPYAP
LBPFOASENGEPMITEAGLNLPVRSADILKALABATELAGELDGVITVPAYE
DLAORGTEDARLAIAGLHVLRLLINEPTLAIVGQIDSOGEVIAYVDLGGGPFISTII
RLSRGFYEVTLARGSDALSGDDPDFDHLDIYREQAGIDPDRSNRYQERLLDAALAAKTI
ALSADSVTVNNAWGDELSTREOFNELIALPVKRTLKCRRLKQAGYBADEVLEVWVA
VGISGIVPLAVREVGEFFPRPLTSIDPKVVALGAILQADILVGNKRKMLLDIAYT
PLSLGLETMGVLGAEKVIPIINTTIPVARADDTFTTFDGOGTAMS IHMOGERSEIMLDCSES

	gene	LARFALRGIPALPAGCAHIRTFOVDADGLLSVTAMKSTSVENSIOVKPSBYGJLDS LAISMLKSDMSVAYEDQVKAKMTAEQVEKEARVDESJHGALAADAALSAREQVIDDAA AHISEVAQGDVDADEOKAIKNVDKQTDFPARRMDQSVRLKHGSVDEV"
CDS	gene	complement(5231..5746) /gene="EC63393" complement(5231..5746) /note="similar to HSB_ECOLI gi 1788876 percent identity 100 in 171 aa (Conserved in E.coli K-12)" /codon_start=1 /evidence=not experimental /transl_table=11 /product="molecular chaperone" /protein_id="BAB36816.1" /db_xref="GI:13362864" /translation="MDYPLFLPYRYOLDTQLSLRFODLRQYHPDKFASGSQAQ LIAVOASATINQMOTLRHLRLRAEYLILSHGFDAISOHTVRDTAFMEOLEBEL DEIEOKADEARLESPIKRVKMFPTRHQLMEOGLNENWDAADIVRKLRFLDKLRSS AEOLSEKLDP" compLement(5842..6165) /gene="EC63394" complement(5842..6165) /gene="EC63394" /note="similar to ISCA_ECOLI gi 1788877 percent identity 100 in 107 aa (Conserved in E.coli K-12)" /codon_start=1 /evidence=not experimental /transl_table=11 /product="putative iron-binding protein" /protein_id="BAB36817.1" /db_xref="GI:13362865" /translation="MSITLSDSAARAVENTFLNRKGFGRLIGVRTSGCSGNAYLEF VSDPTEDIVFDKGVKVVDKSLQFLDGTLDPVEKGLNEGFKFTNPNDCEGC ESPHV"
Query Match	91.7%; Score 687.6; DB 1; Length 270365;	
Best Local Similarity	94.8%; Pred. No. 2.9e-204;	
Matches 711; Conservative	0; Mismatches 39; Indels 0; Gaps 0;	
Oy	1	GATCATTAACCCCAAACCTATCGAATAATTCGAATGTAGAATATAAAAATTTCATTTT 60
Dd	61135	GATCATTAACCCCAAACCTATCGAATAATTCGAATGTAGAATATAAAAATTTCATTTT 61194
Oy	61	TAAATGTTCCGTGTGGGTACTGTCTACCAAAACAGAGATTAACAAGTACACCGAC 120
Dd	61195	TAAATATCTGTGTGGGTACTGTCTACCAAAACAGAGATTAACAAGTACACCGAC 61254
Oy	121	CCTTTAAAGTGTTTTTGACCTTACACCTCGATTACCGETTAGACGCCAGAACCA 180
Dd	61255	CCTTTAAAGTGTTTTTGACCTTACACCTCGATTACCGETTAGACGCCAGAACCA 61314
Oy	181	TATTCGACCCCTTAAGCTCCTCACTGACGATGATGATGATGATGATGATGATGAT 240
Dd	61315	TATTCGACCCCTTAAGCTCCTCACTGACGATGATGATGATGATGATGATGATGAT 61374
Oy	241	AGGATGAGTCTGGATTTTGAATGTGATGTACTGTGTGGGGCATTTTCATTTTCACT 300
Dd	61375	AGGATGAGTCTGGATTTTGAATGTGATGTACTGTGTGGGGCATTTTCATTTTCACT 61434
Oy	301	GGCAGTATTTGACCCGCGACGGGTACACCTTTGAATTTGGGGGGGGGGCGCATATTTC 360
Dd	61435	GGCAGTATTTGACCCGCGACGGGTACACCTTTGAATTTGGGGGGGGGGCGCATATTTC 61494
Oy	361	CTGGCTGGCGGTGAAAATGCGCACAGGCCCAACAAAGAAAGACGAGCTTCAGGCAAAACC 420
Dd	61495	CTGGCTGGCGGTGAAAATGCGCACAGGCCCAACAAAGAAAGACGAGCTTCAGGCAAAACC 61554
Oy	421	AATCAGCTTTTGGCGCAGCTTTGCTTTCGATGATGATGATGATGATGATGATGATGAT 480
Dd	61555	AATCAGCTTTTGGCGCAGCTTTGCTTTCGATGATGATGATGATGATGATGATGATGAT 61614
Oy	481	TGTTAGCGACGTGTGACGTTTGTTCGCGGCAACACAGGCGTTAAGTGGGTATTGG 540

Db 61615 TGTTACGGCACTGTGACATTTGTTCTGCGCAAAACAGCGGCTTAAGCTGATAGTTGG 61674

Qy 541 CGTAGCGCTTTTGTGGGATGATTTGGGACATTTGCTGGGCGCTGGCGGG 600

Db 61675 CGTAGCGCTTTTGTGGGATGATTTGGGACATTTGCTGGGCGCTGGCGGG 61734

Qy 601 GCATCTGTTTACGAGATTTGTTCCGACATATGTCGACGTTAAATATCGTCTGGCCCT 660

Db 61735 GCATCTGTTTACGAGATTTGTTCCGACATATGTCGACGTTAAATATCGTCTGGCCCT 61794

Qy 661 GTTCTGCTCTATTGCGCGGATGTCGATTTTCTATTAAAGAAAAAGCGAAGAGTGGC 720

Db 61795 GCTCTGATCTATTGCGCGGATGTCGATTTTCTATTAAAGAAAAAGCGAAGAGTGGC 61854

Qy 721 CCTTTTCGCTTACTTACTTCTACTTAAG 750

Db 61855 CCTTTTCGCTTACTTACTTCTACTTAAG 61884

RESULT 11

AC126452 194171 bp DNA linear HTG 15-SEP-2004

LOCUS AC126452 Mus musculus chromosome UNKNOWN clone RP23-215C15, WORKING DRAFT

DEFINITION AC126452

ACCESSION AC126452.3 GI:52077982

VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS ACTIVEFIN.

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 194171)

WILSON, R.K.

REFERENCE TITLE The sequence of Mus musculus clone

UNPUBLISHED

2 (bases 1 to 194171)

McPherson, J.D. and Waterston, R.H.

REFERENCE TITLE Direct Submission

AUTHORS Submitted (05-UTL-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

JOURNAL 3 (bases 1 to 194171)

WILSON, R.K.

REFERENCE TITLE Direct Submission

AUTHORS Submitted (15-SEP-2004) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

JOURNAL On Sep 15, 2004 this sequence version replaced gi:38564407.

COMMENT

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu

Contact: submissions@wustl.edu

Project Information -----

Center project name: M BA0215C15

----- Summary Statistics -----

Sequencing vector: M13; 0%

Sequencing vector: Plasmid; 100%

Chemistry: Dye-Primer ET; 0% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 191627 bases at least Q40

Consensus quality: 191807 bases at least Q30

Consensus quality: 191942 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently

* consists of 7 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 1232: contig of 1232 bp in length

* 1233 gap of unknown length

* 1333 contig of 2337 bp in length

* 3669 gap of unknown length

* 3770 contig of 6214 bp in length

* 9984 gap of unknown length

* 10084 contig of 21144 bp in length

* 31228 gap of unknown length

* 31328 contig of 38621 bp in length

* 69949 gap of unknown length

* 70049 gap of unknown length

* 192986 contig of 122938 bp in length

* 192987 gap of unknown length

* 193087 194171: contig of 1085 bp in length.

FEATURES

Location/Qualifiers

Source

1. 194171

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="UNKNOWN"

/clone="RP23-215C15"

1. 1232

/note="assembly_name:Contig42"

1333. 3669

/note="assembly_name:Contig43"

3770. 9983

/note="assembly_name:Contig44"

10084. 31227

/note="assembly_name:Contig45"

31328. 69948

/note="assembly_name:Contig46"

70049. 192986

/note="assembly_name:Contig47"

clone_end:SP6

vector_side:Left"

193087. 194171

/note="assembly_name:Contig7"

ORIGIN

Query Match 83.1%; Score 623.4; DB 2; Length 194171;

Best Local Similarity 99.7%; Pred. No. 4.6e-184;

Matches 635; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 115 ACCGACCCCTTTAAGCTTTTGGACCTTACACCGTATTAACCGTATGACGACGAGC 174

Db 193122 ATGACCCCTTTAAGCTTTTGGACCTTACACCGTATTAACCGTATGACGACGAGC 193181

Qy 175 GAACAATATTCGCGCCCTTACGCTGCTAGCGCATGATTCGTCAAAGTACCGCGT 234

Db 193182 GAACAATATTCGCGCCCTTACGCTGCTAGCGCATGATTCGTCAAAGTACCGCGT 193241

Qy 235 GCTGGCAGGATGATGCTGGGATTTTGAATGATGATTAATGCTGCGGCAATTCATT 294

Db 193242 GCTGGCAGGATGATGCTGGGATTTTGAATGATGATTAATGCTGCGGCAATTCATT 193301

Qy 295 TTACATGCGCATGATTAACCCGACAGCGTACACCTTTGATGCTGGCGGCGGCGATA 354

Db 193302 TTACATGCGCATGATTAACCCGACAGCGTACACCTTTGATGCTGGCGGCGGCGATA 193361

Qy 355 TATTGCTGCGCGCGTGAATAATCGCACCAAGCAAAAGAAAGAGACGACTTCAGCG 414

Db 193362 TATTGCTGCGCGCGTGAATAATCGCACCAAGCAAAAGAAAGAGACGACTTCAGCG 193421

Qy 415 AAAACCAATAGCTTTTGGCGCAGCTTTGCTTGCAGTTTGAAGCTCAAAATCAATTT 474

Db 193422 AAAACCAATAGCTTTTGGCGCAGCTTTGCTTGCAGTTTGAAGCTCAAAATCAATTT 193481

Qy 475 GTAGCGGTAGGACGCTGCGAGTTTGTTCGCGCAAAACAGGCGTTAGCTGGCGT 534

Db 193482 GTAGCGGTAGGACGCTGCGAGTTTGTTCGCGCAAAACAGGCGTTAGCTGGCGT 193541

Qy 535 AGTTGGCGTCAAGCTTTTGGCGATGATGGAAGCTTTGGCAATGCTGCTGGCGCT 594

Db 193542 AGTTGGCGTCAAGCTTTTGGCGATGATGGAAGCTTTGGCAATGCTGCTGGCGCT 193601

QY 595 GGGGGGCGATCTGTTTACGAGTATGTTTCCGACGATAGTCGCGACGTTAAATATCGGCT 654
DB 193602 GGGGGGGGCGATCTGTTTACGAGTATGTTTCCGACGATAGTCGCGACGTTAAATATCGGCT 193661
QY 655 TGCCCTTTGCTGGTCTATTGCGCGGATGCGCATTTTCTATTAAAGAAAAAGCGAGAGA 714
DB 193662 TGCCCTTTGCTGGTCTATTGCGCGGATGCGCATTTTCTATTAAAGAAAAAGCGAGAGA 193721
QY 715 GGTGCGCCCTCTTCG-GCTTAGTAACTTGCTACTTAAG 750
DB 193722 GGTGCGCCCTCTTCGCGGCTTAGTAACTTGCTACTTAAG 193758
RESULT 12
E49391 588 bp DNA linear PAT 31-JAN-2002
LOCUS Process for producing L-amino acid.
DEFINITION E49391
ACCESSION E49391 GI:18628082
VERSION UP 2000189180-A/11.
KEYWORDS Escherichia coli
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 588)
AUTHORS Rivishitsu,V.A., Zakataeva,N.P., Nakanishi,K., Aryoshin,V.V.,
Trosolin,P.V. and Tokmakova,I.R.
TITLE Process for producing L-amino acid
JOURNAL Patent: JP 2000189180-A 11 11-JUL-2000;
AJINOMOTO CO INC
SOURCE Escherichia coli
OS PN JP 2000189180-A/11
PD 11-JUL-2000
PF 28-DEC-1999 JP 1999373651
PR 30-DEC-1998 RU 98124016,09-MAR-1999 RU 99104431 PI
VITARI ARUKAJEVICHI RIVISHITSU,NATARIYA PAVUOVUNA
ZAKATAEVA,
PI KAZUO NAKANISHI,VLADIMIR VENTYAMINOVICH AROSOSHIN, PI PETER
VIRAJUMIROVICHI TOROSHIN,IRINA RIVOVUNA TOKOMAKOVA PC
C12N15/09,C12N1/21,C12P13/04//C12N1/21,C12R1:19), (C12P13/04, PC
C12R1:19),
PC C12N15/00
CC
CDS Key Location/Qualifiers
FH (1)..(588).
FT CDS Location/Qualifiers
1..588
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
ORIGIN
Query Match 78.4%; Score 588; DB 6; Length 588;
Best Local Similarity 100.0%; Pred. No.1.8e-173;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 110 GTGACACGACCCCTTTAAGTCTTTTGGACTTACACCCGATTAACCGCTATGACGCCA 169
DB 1 GTGACACGACCCCTTTAAGTCTTTTGGACTTACACCCGATTAACCGCTATGACGCCA 60
QY 170 GGACCGAACAATATTCGCGCTTAGCTGCTACGTGCGATGAGATTTCGCAAGTACC 229
DB 61 GGACCGAACAATATTCGCGCTTAGCTGCTACGTGCGATGAGATTTCGCAAGTACC 120
QY 230 CGCGTGTGCGAGGATAGTCTGGATTTTGAATTGTGATTAAGTGTGCGGCGCAT 289
DB 121 CGCGTGTGCGAGGATAGTCTGGATTTTGAATTGTGATTAAGTGTGCGGCGCAT 180
QY 290 TCATTTTACGTGCGAGTATGACCCGCGACGGGATACCTTTTGAAGTTGGCGGGGCG 349
DB 181 TCATTTTACGTGCGAGTATGACCCGCGACGGGATACCTTTTGAAGTTGGCGGGGCG 240

QY 350 GCATATATTGTCGCTGGCGTGGAATCGCACGACGCCAACAAGAGACGACCT 409
DB 241 GCATATATTGTCGCTGGCGTGGAATCGCACGACGCCAACAAGAGACGACCT 300
QY 410 CAGGCAAAACCAATGACCTTTTGGGCGACCTTTGCTTGCAGTTGTGAACGTCAAAATC 469
DB 301 CAGGCAAAACCAATGACCTTTTGGGCGACCTTTGCTTGCAGTTGTGAACGTCAAAATC 360
QY 470 ATTTGTAGCGGCTTACGCGACGTGCGAGTTGTTCTCGCGAAACACAGGGCTTAAC 529
DB 361 ATTTGTAGCGGCTTACGCGACGTGCGAGTTGTTCTCGCGAAACACAGGGCTTAAC 420
QY 530 TGGGTAGTTGCGCTCAGCGCTTTTGTGCGAGATGAGACGTTTGGCAATGTGCTG 589
DB 421 TGGGTAGTTGCGCTCAGCGCTTTTGTGCGAGATGAGACGTTTGGCAATGTGCTG 480
QY 590 GCGCTGGCGGGGACATCTGTTTACGACATGTTTCCGACGATGCTGCCAATTAAATATC 649
DB 481 GCGCTGGCGGGGACATCTGTTTACGACATGTTTCCGACGATGCTGCCAATTAAATATC 540
QY 650 GTGCTTGCCTGTTGCTGCTATTTGCGGCGTACGCAATTTCTATTAA 697
DB 541 GTGCTTGCCTGTTGCTGCTATTTGCGGCGTACGCAATTTCTATTAA 588
RESULT 13
AX030083 588 bp DNA linear PAT 16-SEP-2000
LOCUS Sequence 13 from Patent EP1016710.
DEFINITION AX030083
ACCESSION AX030083 GI:10190300
VERSION AX030083.1 GI:10190300
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Nakanishi,K., Alephin,V.V., Liveness,V.A., Tokmakova,I.L.,
Trosolin,P.V. and Zakataeva,N.P.
TITLE Method for producing L-amino acids
JOURNAL Patent: EP 1016710-A 13 05-JUL-2000;
AJINOMOTO KK (JP)
FEATURES
source location/Qualifiers
1..588
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
CDS
1..588
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC09316.1"
/db_xref="GI:10190301"
/translation="MTPTLSAFWYVTLITAMTPGNIIALSSATSHGFRSTRYLA
GMSIGFLIWLICAGISFSLAVIDPAVHLISVAGAAIYVIMLAKIATSPKEDGLOA
KPSFWSFALQFVNKILILGYTALSTFVLPTQGLASVWVGVSVLLAMIGTPGNVCM
ALAGHLFQRLFRQYGRQNLIVALLILVYCAVRFY"
ORIGIN
Query Match 78.4%; Score 588; DB 6; Length 588;
Best Local Similarity 100.0%; Pred. No.1.8e-173;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 110 GTGACACGACCCCTTTAAGTCTTTTGGACTTACACCCGATTAACCGCTATGACGCCA 169
DB 1 GTGACACGACCCCTTTAAGTCTTTTGGACTTACACCCGATTAACCGCTATGACGCCA 60
QY 170 GGACCGAACAATATTCGCGCTTAGCTGCTACGTGCGATGAGATTTCGCAAGTACC 229
DB 61 GGACCGAACAATATTCGCGCTTAGCTGCTACGTGCGATGAGATTTCGCAAGTACC 120
QY 230 CGCGTGTGCGAGGATAGTCTGGATTTTGAATTGTGATTAAGTGTGCGGCGCAT 289

gene /db_xref="UniProt/TREMBL:O8X4S1" /translation="MDWLACWMLIVLVPLVGLVLIYKDKLIDHKPLANKPELP HRDNDKMDDEDMPKQPK" complement (2389..4404) /gene="STY2723" complement (2389..4404) /note="Orthologue of E. coli YPF1_ECOLI; Fasta hit to YPF1_ECOLI (671 aa), 72% identity in 662 aa overlap" /codon_start=1 /transl_table=1 /product="conserved hypothetical protein" /protein_id="CAD02685.1" /db_xref="GI:16503702" /db_xref="GOA:O8Z4S1" /db_xref="UniProt/TREMBL:O8Z4S1" /translation="MSDIDALRALTSQMTGEGIRRLVIGSDAMCEREAIRAVLP GDWVADAPACRPTQALQTLIGREFRAIPDAMGPDAAAFAALSGLOAGSML LIMPVETWESRPDPSLRWSDCAOPITPTPOFOHLKRTLSRDPOTLMRORPCFP SYSPREWRPARGEPDEOAALISRLBEMPGVATVIAPRGRKSLAAGOTISOMGT AITTPAKTADTILAPAGERFCFMAPDALLASGARADMLVDBAATIPAPLLOYS RFPPIILITTVQEGEGTGRGPLKPCARFPOLHRTLRQPVMAPECPLENISEALI FDEBAFOAPHGGAISAFYQOAMGKTPALRAVYQLSGAHYRTSLDIRRMADAG QHFLQATANNRVAGALWLVVEGGISAEISQAVWCGFRRPGNVAOSIAAGSDPLA TLVGRVRSRIAVHAPAROREGIGQOLACACEOAAQCDYLSFGYTELRFMORCGF VIVRMGSRRASSGCTAMALLPLSDAGCLOAHEHRLADRDITLQWNGAIPILA LREQALNDEBERELVGFAPAHRLPLTSLGCHRLQYSLAPLALRGRLEKSKADAEI CARLRISGRKALLQRAQAALIALDAGRTOKLRLVMFGGDHAG" complement (2878..3207) /gene="STY2723" /note="Pfam match to entry PF00583 Acetyltransf. Acetyltransferase (GNAT) family, score 32.50, E-value 9.7e-06" complement (3790..3813) /gene="STY2723" /note="PS00017 ATP/GTP-binding site motif A (P-loop)" complement (4420..5283) /gene="STY2724" complement (4420..5283) /gene="STY2724" /note="Orthologue of E. coli P76563; Fasta hit to P76563 (287 aa), 92% identity in 286 aa overlap. Contains hydrophobic, possible membrane-spanning region" /codon_start=1 /transl_table=1 /product="putative membrane protein" /protein_id="CAD02686.1" /db_xref="GI:16503703" /db_xref="GOA:O8Z4S0" /db_xref="UniProt/TREMBL:O8Z4S0" /translation="MRWQGRRESNNVEDRRNRPGSPSLGSGRPLPRGKSGIILVVY VVAGYGVDTLGLTQGPVSQOSTRSISPDDBAARFSTVILATDITGOLFQKRG RGYOQPLVMYRGMTRTCGAGOSVWGPFCPADGYVILDSFDMKNKLGADAGDQ QGVYIAHEVGHVQKLGIEBKVROLQONMSQTEVNLISVMBLOACCFGWSHSPG QOQVLEAGDLEBALNAOAIGDRLQOQCGRVVDPDSFTGTSBQRYSWPKRGDSGD PAQONTREKKNP" complement (5287..5292) /note="possible RBS" complement (5415..6128) /gene="STY2725" /note="synonym: purC" complement (5415..6128) /gene="STY2725" /note="Orthologue of E. coli purC (PUR7_ECOLI); Fasta hit to PUR7_ECOLI (237 aa), 95% identity in 237 aa overlap" /codon_start=1 /transl_table=1 /product="phosphoribosylaminoimidazole-succinocarboxamide synthase" /protein_id="CAD02687.1" /db_xref="GI:16503704" /db_xref="GOA:O8X69" /db_xref="UniProt/Swiss-Prot:O8X69"

misc_feature /translation="MOKOAELYRGKATVYSTENPDLLVLEPRNDTSAGDARIEQPD RKGCVNNKFNHPIMPTLEAAGIPTQWERLSTPECIYKLEEMVPECVVRNPAAGSLV KVLGESEKNEINPITDLEPLKNDALHDPVNNSSYCTPRMWSQENLAKMKELTYKAND VLKQLFDDAGLILVDFKLEFGLYKGEVVLGDESPDSRLMDKETILDKDFRQSL GGLIEAYEAVHRLGVKLD" complement (5421..6122) /gene="STY2725" /note="Pfam match to entry PF01259 SAICAR synt. SAICAR synthetase, score 414.20, E-value 1.2e-120" complement (5586..5612) /gene="STY2725" /note="PS01058 SAICAR synthetase signature 2" complement (5829..5873) /gene="STY2725"

misc_feature

misc_feature

misc_feature

Query Match 61.5%; Score 461.2; DB 1; Length 145050; Best Local Similarity 78.3%; Pred. No. 5,5e-133; Matches 553; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

1 GATCCATAACCCCAACCTATCGAATAATCGAATCTAGAATATAAATCATTATTTT 60
143748 GATCCATAACCCCAACCTATCGAATAATCTCGAATCTAGAATATAAATCATTATTTT 143807
61 TTAATGTCCTCGTGGGCTACTGTCTACCAAAAAGAGAGATTAACAAGTGAACACGAC 120
143808 TTAATGTCCTCGTGGGCTACTGTCTACCAAAAAGAGAGATTAACAAGTGAACACGAC 143867
121 CTTTAAAGTCTTTTGGACTTAACACCTGATTACCCCTATATAGCCACGACCAACAA 180
143868 GCTTTAAAGCTTTTGGACTTAACACCTGATTACCCCTATATAGCCACGACCAACAA 143927
181 TATTCGCGCCCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
143928 TATTCGCGCCCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 143987
241 AGGATGAGTCTGGGATTTTGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 300
143988 AGGATGAGTCTGGGATTTTGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 144047
301 GGCAGTATGATGACCCCGGACGGTACCTTTGAAGTGGCGGGCGGGCGGATATATTTG 360
144048 TGGCGTATGATGACCCCGGACCTTCAATTTGCTCAGTTGGGTGGGTGGGTGGGTGGGT 144107
361 CTGGCTGGCGGTGAAAAATGCCACAGCCCAACAAAGAAAGAGAGACTTCAGGCAAAACC 420
144108 GTGGCTGGCGGTGAAAAATGCCACAGCTCCAGCGCGGATGAGAAAGTCAAGACAAAACC 144167
421 AATCAGCTTTTGGGCGACGTTTCTTTGCACTTTGCACTTTGCACTTTGCACTTTGCACT 480
144168 GATTCGCTTTTGGGTAAGCTTTGGCGTGGCTTTGGAACGTCAAAATTAATTCGTACCG 144227
481 TGTTCAGGCACTGTCCAGCTTTGTTGTCGCCGAAACACAGCGCTTAAGCTGGTATTTGG 540
144228 CATTCAGGCACTGTCTCAATTTGTTTTCGCAAAAGCAGAGCCCTGAATCGGGTCATTCG 144287
541 CGTCAGGCTTTTGGCGAGATGATGGAGCTTTGGCAATGTGTGCTGGCGCTGGCGGG 600
144288 CGTCAGATATTTGCTGTCATGATCGCACGTTCCGTAATGTGTGCTGGCGCTGGCGGG 144347
601 GCATCTGTTTCAAGAGATTTTTCGCGAGATGATGATGATGATGATGATGATGATGATGAT 660
144348 ACATCTGTTTCAAGAGAGCTTTTGGCAATGATGATGATGATGATGATGATGATGATGAT 144407
661 GTTCTGCTGCTATTTGCGCGGTAGCATTCTTCTTAAAGCAAAAAA 706
144408 ACTGCTGCTGCTGCTGCGCGGTGCGGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 144453

RESULT 15
AE016835/c 300431 bp DNA linear BCT 21-MAR-2003
LOCUS AE016835 Salmonella enterica subsp. enterica serovar Typh1 Ty2, section 2 of 16 of the complete genome.

ACCESSION AE016835 AE014613
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE

AE016835 AE014613
GI:29136420
Salmonella enterica subsp. enterica serovar Typhi Ty2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1 (bases 1 to 300431)
Deng, W., Liou, S.-R., Plunkett, III, G., Mayhew, G. F., Rose, D. J.,
Burland, V., Kodoyianni, V., Schwartz, D. C. and Blattner, F. R.
Comparative Genomics of Salmonella enterica Serovar Typhi Strains
Ty2 and CT18
J. Bacteriol. 185 (7), 2330-2337 (2003)
2 (bases 1 to 300431)
12531367
12644504
Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G. F., Rose, D. J.,
Burland, V., Kodoyianni, V., Schwartz, D. C. and Blattner, F. R.
Direct Submission
Submitted (25-SEP-2002) Laboratory of Genetics, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1. .300431
/organism="Salmonella enterica subsp. enterica serovar
Typhi Ty2"
/mol_type="genomic DNA"
/strain="Ty2"
/serovar="Typhi"
/sub_species="enterica"
/db_xref="taxon:209261"
complement (67. .486)
/gene="yfiG"
/locus_tag="L0261"
complement (67. .486)
/gene="yfiG"
/locus_tag="L0261"
/note="corresponds to STY2842 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=1
/product="chlorodoxin 2"
/protein_id="AA067987.1"
/db_xref="GI:29136421"
/translation="MNTVCTHCOALINRIPGDRLODAKCGHLPDGEVINATGET
LDRLKXDLPIVUIDFMAKMGPCNPAPIEDVABESGKRFKXNTBARELSAFA
GIRSIPEIMIFKIGQVVDMLNAGVPAKAPFSPMLNEAL"
690. .1727
/gene="yfiF"
/locus_tag="L0262"
690. .1727
/gene="yfiF"
/locus_tag="L0262"
/locus_tag="L0262"
/gene="yfiF"
/locus_tag="L0262"
/note="corresponds to STY2841 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=1
/product="putative RNA methyltransferase"
/protein_id="AA067988.1"
/db_xref="GI:29136422"
/translation="MNDLKNKSGKVMYVRSDDSDKRTNHPRTGKGGPAPKASRT
DGRRPAPDRNNOSNRDKHETSPMTVSRAPGDPPEKVDHGSGISFIDEVLR
QARBEFVGENGCALFOSRPAIVAMFIOSTPRFKALRMANKRAYVDEA
ELKASGTENHGCYCPILIKRNGTTCQWTKADDOCVTALEBANFNUHUGMRSK
AHFGVGVVQDALLSGLAIRAEGGAHVOPITGESIVDVLDDFRQAGYVVTTS
SLBGMWRONKA"
complement (1843. .2532)
/gene="ung"
/locus_tag="L0263"
complement (1843. .2532)
/gene="ung"
/locus_tag="L0263"
/locus_tag="L0263"
/note="corresponds to STY2840 from Accession AL513382:

Salmonella typhi CT18"
/codon_start=1
/transl_table=1
/product="uracil-DNA glycosylase"
/protein_id="AA067989.1"
/db_xref="GI:29136423"
/translation="MATELFWHVDLADKQAPFINTLHTVAGROSGITVYPOQXV
FNAFPEIGDVVVLIGDPRVGRQAHLASVPGTAPPSLVNMYEELASIG
FVRPAHGYLSMAROGVLTANTVLTARAQASHASIGMETFTDKYISLJNOREGV
FLMGSHQKKGKAIIDQRRHILKAPHPPLSARHGFPGCNHPLTNQWLEOGEKTI
DMTPVLPABEE"
2851. .3234
/gene="yfiD"
/locus_tag="L0264"
2851. .3234
/gene="yfiD"
/locus_tag="L0264"
/note="corresponds to STY2839 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA067990.1"
/db_xref="GI:29136424"
/translation="MTGIOITKRAANDLINSFMLDSEKGEARCTVAKSGPADEBV
AVSKGIEIRREIEMVEKPRVREGGCHLVNVLKRETTLEDAYKHPKTYQLTIRVSG
YAVRNSLTPEQQRDVARTFTESL"
complement (3296. .3883)
/gene="yfiK"
/locus_tag="L0265"
complement (3296. .3883)
/gene="yfiK"
/locus_tag="L0265"
/locus_tag="L0265"
/note="corresponds to STY2838 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=1
/product="putative membrane protein"
/protein_id="AA067991.1"
/db_xref="GI:29136425"
/translation="MTPMLSAFMYTLITALTGPNNTILASAAVAGFRQIRVLA
GMSLGLVWMLCAGIASLAVIDPAIHLISWGAAYILMKIATSPADEKVP
KPVFMVSRGLQFVNKILLYGTALSTFVLPTQTLNMWVIGVILLALIGTGNVCM
ALAGHLPQARFRYRQGNITIALLVYCAVRLFY"
3986. .4868
/gene="yfiB"
/locus_tag="L0266"
/note="corresponds to STY2837 from Accession AL513382:
Salmonella typhi CT18"
/pseudo
complement (4901. .6235)
/gene="srmb"
/locus_tag="L0267"
complement (4901. .6235)
/gene="srmb"
/locus_tag="L0267"
/note="corresponds to STY2836 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=1
/product="ATP-dependent RNA helicase Srmb"
/protein_id="AA067992.1"
/db_xref="GI:29136426"
/translation="MTVTFSLELDESLLDALQDKGFTRPRAIOAAIPPALDGRDV
LGSAPGTGTAAVYLLPALOHLDPFRKSGSPRIILITPRRLAQVADHAELEAKH
THDIDITIGGVAVNMHAEVSEENODIVATTGRLQYIKENFDCRAVETLLDND
RMLDMGFADIDIEHLAGETRMKOTMLFSATLSDGAIKQFAERLLDPPVVSANSTRB
RKTIHWYTRADNFEHKVALKGLLKQDDATSIYVVRGERRHETLAEITRLAGINNC
YLBGEVAQIKRNGIKRLTDGRVNVAVTDVARGIDIDPVSARVILNFDMPRSDDTLH
KIGRTGRAGRKGTALSLVAHDHLLLKIGTYIEPLKARVIDELAPPTRARBSKLTG
KSKKVLARAEKKEKEKEKEPRVKRYRVDYTNIGKRRRPSGTGKQEOSSSE"
6365. .7102
/gene="yfiC"

CDs

1/locus_tag="t0268"
6365..7102
/gene="yfiC"
/locus_tag="t0268"
/note="corresponds to STY2835 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAO67993.1"
/db_xref="GI:29136427"
/translation="MSQSGSVYRRNGFTFKQPFVAHDRCAKMGVDGILLGAPVAD
VRRIDIGSGSLALMLAQRTDSDVPDAEYELDGAQMAGQAEVAISPWRHRTYHTT
DIDQWAPQVTRFDLIISNPEYTPGVESCTPQEQARYATLIDHQTLLIADCTTT
EDGEFVALPEQIGNAFVQOALNMGMHRLRTRDVAENEARLPHRVLLAFPOADECPSS
DRLVIRGSDQNNSESSTALTOAFYLFM"
complement(7087..8709)
/gene="nadA"
/locus_tag="t0269"
complement(7087..8709)
/gene="nadB"
/locus_tag="t0269"
/note="quinolinate synthetase B; corresponds to STY2834
from Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="L-aspartate oxidase"
/protein_id="AAO67994.1"
/db_xref="GI:29136428"
/translation="MMTTPELSCDVLIIISGAGAGSLALRLAEKHKVIYLSKGVSESG
SFYVAGGIAAFDETDTSIASHEDTLAAGICDRAHVEVVAANAATCYOMLIDGCV
LPTDVQEPGKSYHLTREGGSHRRILLADATKEVETLVSRQNHPIQYTESR
NAVNDIISDKTIGPPRRVVGAMINRNKEWETCHAKSVLATGASKVQYITNDP
ISSGGIAMAWRACGRVANLEFNQHPYALYHPQARNFLTEALRGEGAYLKRPDGSR
FMPDDEYRELAPRDIVARAIIDHEMKRGLGADCMFLDISKEDDFVQHPMIYAKLDD
IGMDLITKEBIPVPAAHYTCGGVVDDYGRDPDGLYVIGESVYGLHGARNMASNSLI
LECLYCGMSAANDIDRRMPSVHSDVLYLAMPBSREYANDERVYIQNHMELRIIMPDY
VGIVATITKLESLRITMLQOEIDEVYIANRVSNNLLLNVLQVALLIVRCAMKERR
ESRGHFTFLDYPQOLAESGPSTLSPLETLINR"

gene

9134..9709
/gene="tpoE"
/locus_tag="t0270"
9134..9709
/gene="tpoE"
/locus_tag="t0270"
/note="sigma-24; corresponds to STY2833 from Accession
AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma-B factor"
/protein_id="AAO67995.1"

CDs

Query Match	61.5%;	Score 461.2;	DB 1;	Length 300431;
Best Local Similarity	78.3%;	Pred. No. 6.4e-133;		
Matches 553;	Conservative 0;	Mismatches 153;	Indels 0;	Gaps 0;

QY	GATCGATTAACCCCAACCTATCGAAATAATGGAATCTAGAAATATATAAAACATTCATTTTT	60
Db	GATCATTAACCAACCAACCTATCGAAATCTGAAATTCGAAATTTAAAAACAATTCATTTTT	3992
QY	TTAAATGTCCTCGTTCGGGTACTGTCTAACAAAACAGAGAGATPAACAAGTAGACCCGAC	61
Db	TTAAATATTTCAATTCGGGTACTGTCTGTATAAACAGAGAGATGACAAAGTAGACCCGAT	3932
QY	CCTTTAAAGTCTTTTGGACTTACACCCGTATTAACGGCTATGACGCCAGAACGAAACA	121
Db	GCCTTTAAGCGCTTTCGAGACTTATACCTATATACCGCCCTTACCCCGGCCAACAACA	3872
QY	TATCTCGCCCTTAAGCTCTGCTAGTCGAGCATGAAATTGCPAAAAGTACCGCGGTCTGGC	181
Db	TATCTCGCGCTTAAGCGCAACCTGCACATAGTGTTTTGGCGACAGATATTCGTGTATCGCC	3812
QY	AGGATAGATCTGGGATTTTGTGATGTGANTTTACTGTGTCGGGCAATTTGATTTTCACT	241
Db	AGGATAGATCTGGGATTTTGTGATGTGANTTTACTGTGTCGGGCAATTTGATTTTCACT	300

Db 3752 AGGATATAGCTGGGAAATTTTGGTCGTCATCTGCTGTGGCCGGGATCCGCTTTTCCT 3693
 Qy 301 GGCAGATATTACCCCGCAGCGGATACACTTTTGAATTGGCCGGGGCGGATATATTTGT 360
 Db 3692 TGGCGTATATGATCCGGCATCATTTCAATTTGTCTCAGTTGGGTAGGGGGCTGCGATATCCT 3633
 Qy 361 CTGGCTGGCCGTGGAAAAATCGCCACACGCCCAAGAAAGACGCACTTCAGGCAAAACC 420
 Db 3632 GTGGCTGCGCGTGGAAAAATTCACACAGTCCAGCGCGCGAATGAAAGGTACAGACAAAAACC 3573
 Qy 421 AATACGCTTTTGGGCCAGCTTTTGCTTTGCAGTTTGTGAAAGTCAAAATCATTTTGTACGG 480
 Db 3572 GGTGGCTTTTGGGTAAAGCTTTTGCCCTGCAGTTTGTGAAGTCAAAATTTTGTGACGG 3513
 Qy 481 TGTTAACGGCACTGTGCAAGTTTGTGTCTGCGCGCAAAACAAGCGCTTAAAGTGGGAGTTGG 540
 Db 3512 CATTAACGCTCTGTCTACATTTTGTTTTAAACCGCAAAACGCAAGCCCTGAATCGGGTCAATCG 3453
 Qy 541 CGTCAGCGCTTTGCTGTGCGATGATTTGGAGCGTTTGGCAATGTGTGCTGGCGCTGGCGGG 600
 Db 3452 CGTCAGATATATGCTGGCATTTGATCCGCAAGCTTCGGTAATGTGTGCTGGCGCGCTGGCGGG 3393
 Qy 601 GCATCTGTTTCAAGGATGTTTTCGCCAGTATGGTCCGCAATTAAATATCGTCTTGCCCT 660
 Db 3392 ACATCTGTTTCAAGGGCGGCGTTTGCCCAATTATGGCGCCAGTTAAATATCATTTGCGGCTT 3333
 Qy 661 GTTGTCTGTCTATTTCCCGCGTACGCATTTTCTATTAACGAAAAAAA 706
 Db 3332 ACTGCTGTGTACTGCGCGGTGGGGAATTTTAACTGTACCAAAATAA 3287

Search completed: September 10, 2005, 08:56:07
Job time : 3399 secs

Job time : 3399 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 10, 2005, 07:40:43 ; Search time 178 Seconds
(without alignments)
6894.424 Million cell updates/sec

Title: US-10-620-487-1

Perfect score: 1 gaccatacaccacaccta.....ttagtaactgtctaactaaag 750

Sequence:

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 81813359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	265.8	35.4	618	4	US-09-489-039A-5490 Sequence 5490, Ap
2	156	20.8	372	4	US-09-489-039A-5497 Sequence 5497, Ap
3	42	5.6	669	4	US-09-328-352-2345 Sequence 2345, Ap
4	39.8	5.3	636	4	US-09-543-681A-2536 Sequence 2536, Ap
5	39.6	5.3	750	4	US-09-252-991A-9146 Sequence 9146, Ap
6	38.2	5.1	636	4	US-09-489-039A-3271 Sequence 3271, Ap
7	35	4.7	46885	4	US-09-949-016-13848 Sequence 13848, A
8	34.2	4.6	645	4	US-09-489-039A-2622 Sequence 2622, Ap
9	33.4	4.5	202	4	US-09-513-999C-29855 Sequence 29855, A
10	33.4	4.5	1055	3	US-09-410-028-1 Sequence 1, Appli
11	33.4	4.5	3077	3	US-09-410-028-3 Sequence 3, Appli
12	33.4	4.5	27056	4	US-09-949-016-12200 Sequence 12200, A
13	33.4	4.5	27057	4	US-09-949-016-13115 Sequence 13115, A
14	33.2	4.4	236964	4	US-09-949-016-15753 Sequence 15753, A
15	33	4.4	72604	3	US-09-268-992-7 Sequence 7, Appli
16	33	4.4	72604	3	US-09-657-474-7 Sequence 7, Appli
17	33	4.4	194933	4	US-09-949-016-14172 Sequence 14172, A
18	32.8	4.4	1194	4	US-09-252-991A-9124 Sequence 9124, A
19	32.8	4.4	1326	4	US-09-252-991A-9206 Sequence 9206, Ap
20	32.8	4.4	1329	4	US-09-252-991A-8853 Sequence 8853, Ap
21	32.8	4.4	234884	4	US-09-949-016-16420 Sequence 16420, A
22	32.6	4.3	505	4	US-09-621-976-15639 Sequence 15639, A
23	32.2	4.3	1001	3	US-09-641-638-340 Sequence 340, App
24	32.2	4.3	1001	4	US-10-170-097-340 Sequence 340, App
25	32.2	4.3	5127	3	US-09-462-606-1 Sequence 1, Appli
26	32.2	4.3	7207	3	US-09-462-606-3 Sequence 3, Appli
27	32.2	4.3	150032	4	US-09-949-016-14321 Sequence 14321, A

C 28	32	4.3	1338	4	US-09-489-039A-6698 Sequence 6698, Ap
C 29	31.8	4.2	789	4	US-09-134-000C-2827 Sequence 2827, Ap
C 30	31.6	4.2	28054	4	US-09-902-540-1248 Sequence 1248, Ap
C 31	31.4	4.2	965	4	US-09-270-767-14185 Sequence 14185, A
C 32	31.4	4.2	197875	4	US-09-949-016-15425 Sequence 15425, A
C 33	31.2	4.2	601	4	US-09-949-016-64764 Sequence 64764, A
C 34	31.2	4.2	601	4	US-09-949-016-205071 Sequence 205071, A
C 35	31.2	4.2	601	4	US-09-949-016-13157 Sequence 13157, A
C 36	31.2	4.2	32010	4	US-09-949-016-17433 Sequence 17433, A
C 37	31.2	4.2	101349	4	US-09-949-016-17539 Sequence 17539, A
C 38	31.2	4.2	218940	4	US-09-949-016-17539 Sequence 17539, A
C 39	31.2	4.2	300402	4	US-09-949-016-13632 Sequence 13632, A
C 40	31.2	4.2	1330025	4	US-09-198-452A-1 Sequence 1, Appli
C 41	31.2	4.2	1230230	4	US-09-438-185A-1 Sequence 1, Appli
C 42	30.8	4.1	726	4	US-09-328-352-3033 Sequence 3033, Ap
C 43	30.8	4.1	936	4	US-09-489-039A-5469 Sequence 5469, Ap
C 44	30.8	4.1	1040	4	US-09-602-787A-305 Sequence 305, App
C 45	30.8	4.1	1497	4	US-09-220-132-94 Sequence 94, Appli

ALIGNMENTS

RESULT 1									
US-09-489-039A-5490									
Sequence 5490, Application US/09489039A									
Patent No. 6610836									
GENERAL INFORMATION:									
APPLICANT: Gary Breton et. al									
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA									
FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS									
CURRENT APPLICATION NUMBER: US/09/489,039A									
CURRENT FILING DATE: 2000-01-27									
PRIOR APPLICATION NUMBER: US 60/117,747									
PRIOR FILING DATE: 1999-01-29									
NUMBER OF SEQ ID NOS: 14342									
SEQ ID NO 5490									
LENGTH: 618									
TYPE: DNA									
ORGANISM: Klebsiella pneumoniae									
US-09-489-039A-5490									
Query Match									
Best Local Similarity 35.4%; Score 265.8; DB 4; Length 618;									
Matches 393; Conservative 0; Mismatches 212; Indels 0; Gaps 0;									
Qy	83	TGTCACCAAAACAGAGATTAACAAGTACACCGACCTTTTAAAGTCTTTTGACT	142						
Db	1	TGTCACCAAAACAGAGATTAACAAGTACACCGACCTTTTAAAGTCTTTTGACT	60						
Qy	143	TACACCTGATTAACCGCTATGACGCGAGACGAAACATATTCGCGCTTAAGTCTGCT	202						
Db	61	TACACCTGATTAACCGCGCTGACACCGCGGCGCAACATATTCGCGCTGAGACGCTC	120						
Qy	203	ACGTCGATGATTTGCTCAAGTACCCGCTGCTGACAGATGATCTGGATTTTGG	262						
Db	121	ACGTCGATGATTTGCTCAAGTACCCGCTGCTGACAGATGATCTGGATTTTGG	180						
Qy	263	ATTGTGATTTTACCTGCTGCGGATTTTCACTTTTCACTGCACTGATTTGACCGGACGC	322						
Db	181	ATTGTGATTTTACCTGCTGCGGATTTTCACTTTTCACTGCACTGATTTGACCGGATTT	240						
Qy	323	GTAACCTTTTGAAGTGGCGGCGGCGGATATATTTCTGCGCTGCGGCGGAAATTCGCC	382						
Db	241	ACCTGATTTTGAAGTGGCGGCGGCGGATATATTTCTGCGCTGCGGCGGAAATTCGCC	300						
Qy	383	ACCAAGCCCAACAAAGAACGACCTTACAGCAAAACCATACGCTTTGGGCGACGTTT	442						
Db	301	AAAAGCAACCGCGGACGAGGACGCGGATGTAACCGGATGCTTTGGGCGAGCTTG	360						
Qy	443	GCTTTCAGTTTGAAGCTCAAAATCATTTTGAAGTGTACGCTGACGACGCTT	502						

Db 361 GGGCTACAGTTGTCACACTTAAATTTATTCGTATGAGCATCAACCGGCTTTCTACCTTT 420

Qy 503 GTTCTGCCGCAACACAGCGCTTAAGCTGGGATAGTTGGCGTCAAGCTTTTGTGCGCATG 562

Db 421 GTTCTGCCGCTTAACCTGCGAGCGCTGTTGGCTTATCAGCGCTGAGTCTGCTGGCGCG 480

Qy 563 ATTGGAGCGTTTGGCAATGTGTGCTGGGCGCTGGCGGGGATCATGTTTCAAGCATTTGTT 622

Db 481 ATTGGCGCGCTGGAAATCTGTGCTGGCGCTGGCGGGGATCATGTTTCAACGCTGTT 540

Qy 623 CGCCAGTATGTGCGCAGTTAAATATTCGTGCTGCCCTGTTGGTGTCTATTGCGCGGTA 682

Db 541 TTGCTGTACGGCGCGCACTGAACGTAGTCTGCTGGCTGTTATTTGTTCTACTGTGCTG 600

Qy 683 CGCAT 687

Db 601 CGCAT 605

RESULT 2

US-09-489-039A-5497

; Sequence 5497, Application US/09489039A

; Patent No. 6610836

GENERAL INFORMATION:

APPLICANT: Gary Breton et. al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709.2004001

CURRENT FILING DATE: 2000-01-27

PRIOR FILING DATE: 1999-01-29

PRIOR APPLICATION NUMBER: US 60/117,747

PRIOR FILING DATE: 1999-01-29

NUMBER OF SEQ ID NOS: 14342

SEQ ID NO 5497

LENGTH: 372

TYPE: DNA

ORGANISM: Klebsiella pneumoniae

US-09-489-039A-5497

Query Match 20.8%; Score 156; DB 4; Length 372;

Best Local Similarity 71.8%; Pred. No. 1.4e-44;

Matches 204; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 19 TATCGAAATATCGAATCTAGAAATATTAATAAATCAATCTTTTAAATGTTCCGTGCG 78

Db 85 TATCGATTTATCGAAGCGAATGAATAAATATTCATTTTTTAAATATACACTTCGG 144

Qy 79 GATCGTCTACCAAAACAGAGGATATACAGTACGACCGACCTTTTAAAGTGGCTTTTG 138

Db 145 GTAGTGTCTGCCAGACAGAGGAGAAACCCTGTACCTTGAATACGCGCTTTT 204

Qy 139 GACTTACACCCGTGATTACCGCTATGACGCGAGACCGAACAATATTCGCTTTAGCTC 198

Db 205 AACTTACACTTTGATGATCAACCGCTGTACACCGGGCCCAACAATATCTCGGGCTGAGAG 264

Qy 199 TGGTACGTGCGATGATTTGTGTCAAAGTACCCGGGTGCTGGACGAGTGAAGTCTGGGATT 258

Db 265 CGTACCTCCCATGCTCTACGCGCGACCTGCGGGTGTCTGGCCGGAGTGAAGCTGCGGTT 324

Qy 259 TTTGATTTGATGTTACTGTGTGCGGGCATTTTCAATTTTCACTGG 302

Db 325 TATTATCACTATGTTGATTTGGCGCGGCTTTTAACTTTTCGTGG 368

RESULT 3

US-09-328-352-2345

; Sequence 2345, Application US/09328352

; Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

TITLE OF INVENTION: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC99-03PA

```

CURRENT APPLICATION NUMBER: US/09/328,352
;
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2345
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2345

Query Match      5.6%; Score 42; DB 4; Length 669;
Best Local Similarity 50.4%; Pred. No. 0.0056;
Matches 133; Conservative 0; Mismatches 125; Indels 6; Gaps 1;

QY 105 AACAGTACACCCCTTTAGTCTTTTGACATACACCTGATTACCGCTAAGA 164
DB 53 AGCATATGAGCTTTTAACTTTTATGATTTTGAAGTCTCAATCTTTTATTATA 112
QY 165 CGCCAGAGCCGACATATTTCTCGCCCTTAGCTCTGCTACGTCGATGATTTGTCMAA 224
DB 113 CGCCAGAGGCGAGATCGGCGCCATTCGCCGATTTAAAGGTAAAGT-----CG 166
QY 225 GTACCCCGGCTGCTGCGAGGAGATGAGTCTGGGATTTTGTATGTATCTGTGCGG 284
DB 167 TCGTCCCGGCTGTGACAGTATGCTATTCGGGCACTTATTAAGATTTATGGTAGCGG 226
QY 285 GCATTTTCACTTCTGCTGCAATGATGACCCGCGAGGGATACACTTTTGAGTTGGCGG 344
DB 227 CTGGTGTGCTCTGCTGTAGCAATAACCACTGCTGATGATTTCTTACTGTGCGG 286
QY 345 GGGGCGCATATATTGTCTGCGCTGG 368
DB 287 GTTCTGCTTATTTATATGATGG 310

RESULT 4
US-09-543-681A-2536
; Sequence 2536, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2536
; LENGTH: 636
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2536

Query Match      5.3%; Score 39.8; DB 4; Length 636;
Best Local Similarity 46.5%; Pred. No. 0.0033;
Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 111 TGACACGACCCCTTTAAGTCTTTTGGACCTTACACCCGATTAACGCTATGACGCCAG 170
DB 32 TGACAAATAGCTTACTAGTTGTTTCGTTGACGCTTTTCCTTTATTTGACGCAATTACGCCG 91
QY 171 GACCGACATATTTCTCGCCCTTAGCTCTGCTAGCGTACGTCGATGATTTGCTGAAATACCC 230
DB 92 GACCTAATAATTTACTATTAATACCTTTGGGGGGCTCAGTTGGTTTAAAGCTCTTTAA 151
QY 231 GCGTGTGCGCAGGAGTGTCTGGGATTTTGTATTTGATGTACTGTGTGCGGGCATTT 290
DB 152 CGTTGATGCGAGGCAATTATCGTCGGAATGACATGTATTAATTAGCTCTTCGTTTGTA 211
QY 291 CATTTTCACTGAGCATGATGACCCGCGAGGGATACACTTTTGAGTTGGGCGGGGCGG 350
DB 212 TTGGCAGCTTATTGATTTATTTATCTGCAATTCGATTCGATTCGTTTAAAGATTTATGGTAGCA 271

```

```

QY      351 CATATATGTCCTGGCTGGCCGCGGAAAATCGCCACC 385
          |||  |||  |||  |||  |||  |||  |||  |||
Db      272 TTATCTTGTGGTTAGCTTGAAAAAGACAACC 306

```

RESULT 5
US-09-252-991A-9146

```

: Sequence 9146, Application US/09252991A
: Patent No. 6551795
:
: GENERAL INFORMATION:
:   APPLICANT: Marc J. Rubenfield et al.
:   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS.
:   TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
:   FILE REFERENCE: 107196.136
:   CURRENT APPLICATION NUMBER: US/09/252,991A
:   CURRENT FILING DATE: 1999-02-18
:   PRIOR APPLICATION NUMBER: US 60/074,788
:   PRIOR FILING DATE: 1998-02-18
:   PRIOR APPLICATION NUMBER: US 60/094,190
:   PRIOR FILING DATE: 1998-07-27
:   NUMBER OF SEQ ID NOS: 33142
:   SEQ ID NO 9146
:   LENGTH: 750
:   TYPE: DNA
:   ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-9146

```

Query Match	5.3%	Score 39.6	DB 4	Length 750
Best Local Similarity	47.2%	Pred. No. 0.0043		
Matches 120, Conservative	0	Mismatches 134	Indels 0	Gaps 0

Qy	131	GCCTTTGGACTTAACACCCTGATTAACCGGTATGAGCCAGACCGAACAATATTTCTGGCC	190
Db	160	GGCTTCTCCCTGTTCCGCTTTCGCACTTCGTCAGCCGCGGTCCGAAACAACATGTTG	219
Qy	191	CTTAAGCTGTGCTACGTGCGATGGATTTGCTCAAAGTACCCGCGTGTGCGAGGATGAGT	250
Db	220	CTGAGCTCGGGGGGTCAACTTCGGGTTTGTGCCCTCCATCCGACATCTCGGGATCAGT	279
Qy	251	CTGGAGATTTTGTATGTGATGATGTACTGTGTGGGGCATTTTCAATTTCTAGCGAGTATTT	310
Db	280	TGCGGCTTCTTCAATCATGGTGATGGCGGTGGGCTTCGCTGAGGACAGGTGTTTGAGGCC	339
Qy	311	GACCCGCGACGGGTACACCTTTTGAATTTGGGGGGGGCGGACATATATTTGTCTGGCTGGCG	370
Db	340	TATCCGGTGTGTACACCATCTCGCGTACGTGGGGGGGCTTACCTCTGTACTCTGGCC	399
Qy	371	TGGAATATCGCCAC	384
Db	400	TGAGAGTCCGCAC	413

RESULT 6

```

US-09-489-039A-3271
: Sequence 3271, Application US/09489039A
: Patent No. 6610836
: GENERAL INFORMATION:
: APPLICANT: Gary Breton et. al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
: TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 2709.2004001
: CURRENT APPLICATION NUMBER: US/09/489,039A
: CURRENT FILING DATE: 2000-01-27
: PRIOR APPLICATION NUMBER: US 60/117,747
: PRIOR FILING DATE: 1999-01-29
: NUMBER OF SEQ ID NOS: 14342
: SEQ ID NO 3271
: LENGTH: 636
: TYPE: DNA
: ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3271

```

Query Match	5.1%;	Score 38.2;	DB 4;	Length 636;
Best Local Similarity	45.9%;	Pred. No. 0.012;		
Matches 130; Conservative	0;	Mismatches 153;	Indels 0;	Gaps 0;

Qy	102	GATACAAAGTACACCGACCCCTTTAAAGTGTCTTTTGGACTTACCAACCCGTATACCGCTA	161
Db	29	GATATTTTCATCTCAACGAATTTCTGATCGCATTTATTTATTTATGCAATTTGTGTAACCTCA	88
Qy	162	TGACGCCAGACCGAACAATTTCTCGCCCTTACCTCTGTACGTCGATGATTTGCTC	221
Db	89	TTACACCCCGGCTCAATAACACTATGTTGCTGTGATCAGGTCTGAATATATGTTTTAAAC	148
Qy	222	AAAGTACCCCGGTGCTGGCAGAGATGAGTGGGATTTTTCATTTGTAATGTTTACTGTGTG	281
Db	149	GTAAGCTTCCCAATGTTCTTGGCATTCAGACATTTGGTTTGTCTCATATGTCATGGCCGTG	208
Qy	282	CGGGCAATTCATTTTCACTGGCAGTGAATGACCCGGCAGCGGTACACCTTTTGAAGTTGGG	341
Db	209	GTACTGTGCTGTGTCGTGTGTTTAAAAATTTGGCCTCAGTGTATATATAGCCTTACGATAT	268
Qy	342	CGGGGGGGGACATATATTGTCTGGGTGGGTGGGAAATGGCCAC	384
Db	269	GGGTGGACTTATTTATATATATCTGGCATGGAATATAGCCAC	311

RESULT 7

```

US-09-949-016-13848/c
; Sequence 13848, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13848
; LENGTH: 46885
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1) ..(46885)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13848

```

Query Match 4.7%; Score 35; DB 4; Length 46885;

Best Local Similarity 56.1%; Pred.No. 3.1;
Matches 87; Conservative 0; Mismatches 65; Indels 3; Gaps 1.

Qy	3	TCGATAACCCCAAAACCTATGCAGAAATATTCAGAAATTTAAATCAATTCATTTTT	62
Db	2108	TCCTAAACCAAAACAGAGAAAAAATATCCACAGAAATCAAAAACCTGATTTTCT	2049
Qy	63	AAATGTTCCGT---GTGGGTACTGTCTTACCAAAACAGAGAGATTAACAGTGACACGA	119
Db	2048	CCCTCTCTGTCTGTGCTCTTGTAAATGGCTACCTGGAGACATCAGAGACAGATTTAAAGCA	1988
Qy	120	CCCTTTTAAGTGTCTTTTGGACTTACCCCTGATT	154
Db	1988	ATCTTTTTTATCATGAGGTACTTACAGAGTAATT	1954

RESULT 8

US-09-489-039A-2622

```
; Sequence 2622, Application US/09489039A
; Patient No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2622
; LENGTH: 645
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-2622

Query Match          4.6%; Score 34.2; DB 4; Length 645;
Best Local Similarity 47.8%; Pred. No. 0.32;
Matches 99; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY      162 TGACGCCAGACGACAAATATTCGCGCTTACCTCTGCTACGTCGATGATTTGCTC 221
DB      80 TGACGCCAGGCTTGATATCTGCTCTTATTCCTCCGACGCGACATGCTGAAGGGGACGCA 139

QY      222 AAAGTACCCGCTGCTGCGAGGATGAGTCTGGGATTTTGATTTGATGATGTTACTGCTG 281
DB      140 AAGCTTGATGCGCGCTTGAGCATGCTCGGGTCTTATCTGGGCGCGCTGCTGG 199

QY      282 CGGCGATTTGATTTTCACTGGCAGTGAATTGACCCGCGACGCGTACACCTTTTGAGTTGGG 341
DB      200 CCTTTGGTTTAGGTAGCGCTGCTACTGCGGCTTCTGAGTTGGCTTATACCTTATGAAATG 259

QY      342 CGGGGCGCGCATATATTTGCTGCTGG 368
DB      260 GCGAGCAGCATACCTTTGCTGCTGG 286

RESULT 9
US-09-513-999C-29855
; Sequence 29855, Application US/09513999C
; Patient No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59, US2, REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 29855
; LENGTH: 202
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 173
; OTHER INFORMATION: m=a or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 189
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 197
; OTHER INFORMATION: m=a or c
; US-09-513-999C-29855
```

```
Query Match          4.5%; Score 33.4; DB 4; Length 202;
Best Local Similarity 54.0%; Pred. No. 0.28;
Matches 61; Conservative 3; Mismatches 49; Indels 0; Gaps 0;

QY      3 TCCATACCCCAACCTATCGAAATATCGAATCTAGAAATATATTAACCTTATTTT 62
DB      86 TACATTAAGTAATATCTTGAATCTTAATTAATTCAGATTTAATTAATATATATTTT 145

QY      63 AATGTTCCGTGCGGCTGCTGCTACCAAAACAGAGATATACAGTACA 115
DB      146 ATATCTTTGTTGATTAATAATGTTTAAACCTAAAAATTAARCATTTGAAA 198

RESULT 10
US-09-410-028-1/c
; Sequence 1, Application US/09410028
; Patient No. 6107471
; GENERAL INFORMATION:
; APPLICANT: ARAHIRA, Masaomi
; APPLICANT: FUKAZAWA, Chikafusa
; TITLE OF INVENTION: PLANT-DERIVED, ASPARAGINE RESIDUE-SPECIFIC ENDOPROTEASE
; FILE REFERENCE: 8361-0007-0
; CURRENT APPLICATION NUMBER: US/09/410,028
; CURRENT FILING DATE: 1999-10-01
; EARLIER APPLICATION NUMBER: JP 327537/1998
; EARLIER FILING DATE: 1998-11-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1055
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (85)..(810)
; US-09-410-028-1

Query Match          4.5%; Score 33.4; DB 3; Length 1055;
Best Local Similarity 55.7%; Pred. No. 0.86;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY      496 GACGTTGTTCTGCGCGAAACACAGCGTTAACTGGTAGTTGGCGTCAGCGTTTCT 555
DB      294 GAGTTTGGCATTTGCCCAAGAAAGAGATTGAGGTGAAGATTTGGGGCAAGATCTGGGT 235

QY      556 GCGATGATTTGGAGCGTTTGGCAATGTGTGCTGGCGGCTGGCGGCGATCTGTT 610
DB      234 GAAGATTTTGGGAGGTGATTCAGATGAAGTTGGCGGCTGTAGGGCGCATTAAGCTT 180

RESULT 11
US-09-410-028-3/c
; Sequence 3, Application US/09410028
; Patient No. 6107471
; GENERAL INFORMATION:
; APPLICANT: ARAHIRA, Masaomi
; APPLICANT: FUKAZAWA, Chikafusa
; TITLE OF INVENTION: PLANT-DERIVED, ASPARAGINE RESIDUE-SPECIFIC ENDOPROTEASE
; FILE REFERENCE: 8361-0007-0
; CURRENT APPLICATION NUMBER: US/09/410,028
; CURRENT FILING DATE: 1999-10-01
; EARLIER APPLICATION NUMBER: JP 327537/1998
; EARLIER FILING DATE: 1998-11-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3077
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: (85)..(321)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (389)..(454)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1160)..(1300)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2414)..(2506)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2707)..(2895)
US-09-410-028-3
```

```
Query Match 4.5%; Score 33.4; DB 3; Length 3077;
Best Local Similarity 55.7%; Pred. No. 1.8;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
```

```
QY 496 GACGTTGTCCTGCGGAAACACAGGCGTTAAGCTGGGTAGTGGCGTCAAGCTTTGCT 555
DB 294 GAGTTGGCAATTTGCCCCAGAGAGAGAGTTGAGGTGAACAGATTGGGGCAAGATCTGGGT 235
QY 556 GCGGATATTGGGACGTTTGGCAATGTGTGCGGGCGCTGGCGGGGCAATCTGTT 610
DB 234 GAAGATTTTGGAGGATGATTCACGATGAAGTTGGCCCTGTAGGGGCAATGACTT 180
```

```
RESULT 12
US-09-949-016-12200
; Sequence 12200, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12200
; LENGTH: 27056
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12200
```

```
Query Match 4.5%; Score 33.4; DB 4; Length 27056;
Best Local Similarity 58.6%; Pred. No. 7.9;
Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```
QY 537 TTGGCGTACAGCTTTTCTGCGCAGATGATGGAGCGTTGGCAATGTGCTGGCGCTGG 596
DB 3385 TTTGTGTCTGCGTGTACCGGAGAGCGGGTGGAGGGGTGCAAGTGTGATGTGGGGG 3444
QY 597 CCGGGCATCTGTTTACAGCAATGTTTGGCCAGTATGCTC 635
DB 3445 GGGGGGATATGTCTGGAGGCTGCCGTCCAGGAGGCTC 3483
```

```
RESULT 13
US-09-949-016-13115
; Sequence 13115, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13115
; LENGTH: 27057
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13115
```

```
Query Match 4.5%; Score 33.4; DB 4; Length 27057;
Best Local Similarity 58.6%; Pred. No. 7.9;
Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```
QY 537 TTGGCGTACAGCTTTTCTGCGCAGATGATGGAGCGTTTGGCAATGTGCTGGCGCTGG 596
DB 3385 TTTGTGTCTGCGTGTACCGGAGAGCGGGTGGAGGGGTGCAAGTGTGATGTGGGGG 3444
QY 597 CCGGGCATCTGTTTACAGCAATGTTTGGCCAGTATGCTC 635
DB 3445 GGGGGGATATGTCTGGAGGCTGCCGTCCAGGAGGCTC 3483
```

```
RESULT 14
US-09-949-016-15753
; Sequence 15753, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15753
; LENGTH: 236964
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)..(236964)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15753
```

```
Query Match 4.4%; Score 33.2; DB 4; Length 236964;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
```

```
QY 353 TATATGTCTGCTGCGTGGAAATGCGCACAGCCCAAGAGAGAGCACTTTCAG 412
DB 234821 TATAGCTTTATACGCTGTGCGCATATTAATGTTTAAATAGCAAGAGCAAT 234880
QY 413 GCAAAACCAATCAGCTTTTGGCGCACAGCTTTGCTTGCAGTTGTGAACGTCAAAATCATT 472
DB 234881 CAAAACCTTTTGTGTTGTAGCTGCTTTTGCCTTTCACAAATTTGCTTACAAATGTTTT 234940
```

QY 473 TTGTACGGTGTACGGCACTGCGACCTGTGCTCGCGCAACAC 518
| | | | | | | | | | | | | | | | | | | | | |
Db 234941 TGTTAATAATAATATGACCTGTGTATCTGTACTGTGCAATAC 234986

RESULT 15

US-09-268-992-7
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 72604
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified base
; LOCATION: all n positions
; OTHER INFORMATION: n=a, c, g, or t
US-09-268-992-7

Query Match 4.4%; Score 33; DB 3; Length 72604;

Best Local Similarity 50.3%; Pred. No. 21;

Matches 81; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 47 AACATTCAATTTTAAATGTTCCGTGCGGTAAGTCTTACCAAAACAGAGAGATAA 106
| | | | | | | | | | | | | | | | | | | | | |
Db 22828 ATACATATATTTATATATACATACACACACTTTATTAATACCAATAGTATCTTTT 22887
| | | | | | | | | | | | | | | | | | | | | |
QY 107 CAATGACACCGACCTTTTAAGTCTTTTGAAGTACACCTGATTACCGTATGACG 166
| | | | | | | | | | | | | | | | | | | | | |
Db 22888 CAAATGAAACAGTAATTTAACAATACTATGAACCTTAATCTAAAGTAAAACTGACA 22947
| | | | | | | | | | | | | | | | | | | | | |
QY 167 CCAGGACCGAACAATATTTCTGCCCTTACGCTCTGCTACGTC 207
| | | | | | | | | | | | | | | | | | | | | |
Db 22948 ACAGTGATGCGAATTTTGTCTCTTACGCTCAAGTACGTC 22988
| | | | | | | | | | | | | | | | | | | | | |

Search completed: September 10, 2005, 09:50:53
Job time : 181 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 10, 2005, 08:56:14 ; Search time 625 Seconds

(without alignments)
7880.689 Million cell updates/sec

Title: US-10-620-487-1

Perfect score: 750
Sequence: 1 gatccataaccacaaccta.....ttagtaactgtcctaaga 750

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7351250 seqs, 3283620254 residues

Total number of hits satisfying chosen parameters: 14702500

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/prodata/1/pubpna/PC07_NEW_PUB.seq:*

3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/prodata/1/pubpna/PC08_PUBCOMB.seq:*

7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*

10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*

11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*

13: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*

14: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*

15: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*

16: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*

17: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*

18: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*

19: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*

20: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*

21: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*

22: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*

23: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*

24: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*

25: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*

26: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	750	100.0	750	18	US-10-620-487-1
2	50.4	6.7	573	9	US-09-974-100-434
3	39	5.2	247682	18	US-10-235-192A-28
4	36.8	4.9	852	18	US-10-425-114-3596
5	36.8	4.9	2727	20	US-10-425-115-65647
6	36	4.8	486	19	US-10-767-795-2906
7	35.2	4.7	474	17	US-10-259-194A-119

Query	Match	Score	DB	Length	ID	Description
1	100.0%	750	18	750	US-10-620-487-1	Sequence 223, App
2	100.0%	750	18	750	US-10-620-487-1	Sequence 681, App
3	100.0%	750	18	750	US-10-620-487-1	Sequence 20855, A
4	100.0%	750	18	750	US-10-620-487-1	Sequence 25623, A
5	100.0%	750	18	750	US-10-620-487-1	Sequence 75212, A
6	100.0%	750	18	750	US-10-620-487-1	Sequence 22, App
7	100.0%	750	18	750	US-10-620-487-1	Sequence 84060, A
8	100.0%	750	18	750	US-10-620-487-1	Sequence 28594, A
9	100.0%	750	18	750	US-10-620-487-1	Sequence 10, App
10	100.0%	750	18	750	US-10-620-487-1	Sequence 17, App
11	100.0%	750	18	750	US-10-620-487-1	Sequence 466, App
12	100.0%	750	18	750	US-10-620-487-1	Sequence 141309, A
13	100.0%	750	18	750	US-10-620-487-1	Sequence 9789, A
14	100.0%	750	18	750	US-10-620-487-1	Sequence 7, App
15	100.0%	750	18	750	US-10-620-487-1	Sequence 1, App
16	100.0%	750	18	750	US-10-620-487-1	Sequence 6876, App
17	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
18	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
19	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
20	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
21	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
22	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
23	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
24	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
25	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
26	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
27	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
28	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
29	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
30	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
31	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
32	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
33	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
34	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
35	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
36	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
37	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
38	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
39	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
40	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
41	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
42	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
43	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
44	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App
45	100.0%	750	18	750	US-10-620-487-1	Sequence 126, App

ALIGNMENTS

RESULT 1

US-10-620-487-1

Sequence 1, Application US/10620487

Publication No. US20040038352A1

GENERAL INFORMATION:

APPLICANT: MAIER, THOMAS

TITLE OF INVENTION: METHOD FOR FERMENTATIVE PRODUCTION OF AMINO ACIDS AND AMINO ACID

FILE REFERENCE: MAIER, T-2

CURRENT APPLICATION NUMBER: US/10/620,487

PRIOR FILING DATE: 2003-07-16

PRIOR APPLICATION NUMBER: GERMAN NO. 102 32 930.3

NUMBER OF SEQ ID NOS: 4

SOFTWARE: PatentIn version 3.1

SEQ ID NO 1

LENGTH: 750

TYPE: DNA

ORGANISM: Escherichia coli

FEATURE:

NAME/KEY: CDS

LOCATION: (110) ..(694)

OTHER INFORMATION:

US-10-620-487-1

Query Match

Best Local Similarity 100.0%; Score 750; DB 18; Length 750;

Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GATCCATAACCCCAACCTATGCAATATCTGATCTAATATTAATAAACATTGATTTT 60

```

Db      1 GATCCATACCCCAACCTATCGAAATATCGAATCTAGAAATATAAAAATCATTTT 60
Qy      61 TTAATATTCCTGTCGGGTACTGTCTACCAAAAACAGAGATTAACAGTACCCGAC 120
Db      61 TTAATATTCCTGTCGGGTACTGTCTACCAAAAACAGAGATTAACAGTACCCGAC 120
Qy      121 CCTTTAAGTCTTTTGGACCTTACACCGTATGACCGGACAGACCA 180
Db      121 CCTTTAAGTCTTTTGGACCTTACACCGTATGACCGGACAGACCA 180
Qy      181 TATTCGCGCCCTTACCTCTGCTACGTCGCAATGATTCGTCAAAATACCCGCTGCGC 240
Db      181 TATTCGCGCCCTTACCTCTGCTACGTCGCAATGATTCGTCAAAATACCCGCTGCGC 240
Qy      241 AGGATAGTCTGGGATTTTGTATGATGATTACTGTGTGCGGCAATTTCAATTTTACT 300
Db      241 AGGATAGTCTGGGATTTTGTATGATGATTACTGTGTGCGGCAATTTCAATTTTACT 300
Qy      301 GGCAATGATGACCCGCGACGCGTACACCTTTTGAATGGGCGGCGGCAATTTTACT 360
Db      301 GGCAATGATGACCCGCGACGCGTACACCTTTTGAATGGGCGGCGGCAATTTTACT 360
Qy      361 CTGCGTGGCGTGAATAATCGCCACACGCCCCAAGAGAGACGACTTCAGGCAAAACC 420
Db      361 CTGCGTGGCGTGAATAATCGCCACACGCCCCAAGAGAGACGACTTCAGGCAAAACC 420
Qy      421 AATAGCTTTTGGGCGCAGCTTTGCTTGGCAATTTGTGAACGTCAAAATCAATTTTGTACGG 480
Db      421 AATAGCTTTTGGGCGCAGCTTTGCTTGGCAATTTGTGAACGTCAAAATCAATTTTGTACGG 480
Qy      481 TGTACGCGACCTGTCGACGTTTGTCTGCGCCAAACACAGCGCTTAAGCTGGGTAGTTGG 540
Db      481 TGTACGCGACCTGTCGACGTTTGTCTGCGCCAAACACAGCGCTTAAGCTGGGTAGTTGG 540
Qy      541 CGTCAGCGTTTGTCTGCGATGATTTGGACGTTTGGCAATGTGTCTGGGCGCTGGCGGG 600
Db      541 CGTCAGCGTTTGTCTGCGATGATTTGGACGTTTGGCAATGTGTCTGGGCGCTGGCGGG 600
Qy      601 GCATCTGTTTACGAGATTTTGGCCAGATAGTGTCCGCAATTAATTCGTGCTTGGCCCT 660
Db      601 GCATCTGTTTACGAGATTTTGGCCAGATAGTGTCCGCAATTAATTCGTGCTTGGCCCT 660
Qy      661 GTTCTGCTATTTGGCGCGGTACCATTTTCTATTAACGAAAAAAGCGAAGAGTCTGC 720
Db      661 GTTCTGCTATTTGGCGCGGTACCATTTTCTATTAACGAAAAAAGCGAAGAGTCTGC 720
Qy      721 CCTCTTCGCTTAACTTGAATTTGCTACTTAAG 750
Db      721 CCTCTTCGCTTAACTTGAATTTGCTACTTAAG 750

```

```

RESULT 2
US-09-974-300-434
; Sequence 434, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Betka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085, 500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 434
; LENGTH: 573
; TYPE: DNA

```

```

; ORGANISM: Bacillus licheniformis
US-09-974-300-434
Query Match      6.7%; Score 50.4; DB 9; Length 573;
Best Local Similarity 46.6%; Pred. No. 1,5e-05;
Matches 194; Conservative 0; Mismatches 221; Indels 1; Gaps 1;

Qy      124 TTTAAGTCTTTTGGACCTTACACCCGTATGACCGGACAGACCGAACAATAT 183
Db      6  TATTAAGCTTTTCTTTCGTATGTATGATCATCATTAAGCCGGGCGGAGCAATAT 65
Qy      184 TCTGCGCCCTTACCTCTGCTACGTCGATGATTCGTCAAAATACCCGCTGCGCAG 243
Db      66  TTTAATGATGAATGAAGCCCAAGGTTCCGCTTTACAGGTTACGCTTTTACGACCG 125
Qy      244 GATGATCTGGGATTTTGTATGATGATTACTGTGTGCGGCAATTTCAATTTTACTGCG 303
Db      126 TATCTTGGCGGGTTTGGACGATCTGGGATTCACGCGGTGCTTACATCAGCTGTGA 185
Qy      304 AGTATTTGACCCGCGACGCGTACACCTTTTGAATGGGCGGCGGCAATATTTGTCTG 363
Db      186 CATTTGGATTCCTGTTGTAGAGCTTATTTAACTTGCAGGCGCTGTATTTGATTTTA 245
Qy      364 GCTGCGGTGAATAATCGCCACACGCCCCAAGAGAGACGACTTCAGGCAAAACCAT 423
Db      246 TTTGGCTTTGCAAGTCGGCTTTTACCAAAATATAAAGAGATTCACAGAAAGCCGCTC 305
Qy      424 CAGCTTTTGGGCGCAGCTTTTGTGCAATTTGTGAACGTCAAAATCAATTTTGTAGCGTGT 483
Db      306 CTCTTTTATATTCGGCTTTTATATTTACGCTGATCAATATTTAAAGCATTTTGTCTTCAT 365
Qy      484 T-ACGCGACCTGTCGACGTTTGTCTGCGCCAAACACAGCGCTTAACTGGGTAGTT 538
Db      366 TAACTGATGACCGGCTTGTCTTTCGCTTCAACATTCCTGAATGACAAATT 421

RESULT 3
US-10-235-192A-28
; Sequence 28, Application US/10235192A
; Publication No. US2004004389A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; TITLE OF INVENTION: Methods and Compositions for Identifying
; TITLE OF INVENTION: Risk Factors for Abnormal Lipid Levels and the Diseases
; TITLE OF INVENTION: and Disorders Associated Therewith
; FILE REFERENCE: MMI-011
; CURRENT APPLICATION NUMBER: US/10/235,192A
; CURRENT FILING DATE: 2002-09-04
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 247682
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 145329-145428, 187896-187995, 195894-195993
; OTHER INFORMATION: N = any nucleotide
US-10-235-192A-28

Query Match      5.2%; Score 39; DB 18; Length 247682;
Best Local Similarity 55.6%; Pred. No. 2.9;
Matches 75; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy      6  ATAAACCCCAACTATCGAAATATGGAATCTGAAATATTAACATTTATTTTAAAA 65
Db      168128 ATAAACCTATATCTATCAATAATTAATGAATTAATTAATTAACCTTCACAAAGAAA 168187
Qy      66  TGTTCGCTGTGCGGTACTGTCTACCAAAAACAGAGGATTAACAGTGAACGACCGCTTT 125
Db      168188 ATGCCAGTCCCAAGATGATTAATTTATCAAAAATGTAAGAGAAATTAACCTAATCTCAT 168247
Qy      126 TAAAGTCTTTTGA 140

```

US-10-425-115-65647

Query Match	4.9%;	Score 36.8;	DB 20;	Length 2727;
-------------	-------	-------------	--------	--------------

Matches	92;	Conservative	0;	Mismatches	92;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

281 GCGGCAATTCATTTCACTGGCAGTGTGACCCGGCAGCGGTACACCTTTTGAGTTGG 340

Db 1741 GCAGCCAGCCAA1GGTCCGTGACCGTGA1GCACCGAGGAGCGTGCAT1GGAGCGAGCTC 1800

341 GCGGGGCGCAATAATGTCCTGGCTGGCGTGGAAATCGCCACGCGCAACCAAGGA 400

DB 1801 AGAGCAGC CGCAGTAT TGTGTG CAGTGT GTCACACAGC GCGTGC TATCAT CACACAGC CAGC 1800

[illegible]

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2818

— — — — —

RESULT 6
WE 10 777 707 0000/2

Sequence 2906, Application US/10767795
Publication No. US2006/0181820A1

```
; GENERAL INFORMATION:
: APPLICANT: Kovalic David R
```

APPLICANT: Cao, Yongwei
APPLICANT: Zhou, Yihua

1; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement

```

; FILE REFERENCE: 38-21(53534)B
; CURRENT APPLICATION NUMBER: US/10/767,795

```

; CURRENT FILING DATE: 2004-01-30
 ; NUMBER OF SEQ ID NOS: 117596

```

; SEQ ID NO 2500
; LENGTH: 486

```

ORGANISM: *Gossypium hirsutum*

NAME/KEY: unsure

OTHER INFORMATION: unsure at all n locations

OTHER INFORMATION: Clone ID: uc-gsflmaxxa076f07b1
; HIS 10 767 705 2006

Query Match 4 84: Score 36: DB 19: Length 486:

Best Local Similarity 52.7%; Pred. No. 0.86;
Matches 78: Conservative 0: Mismatches 70: Indels 0: Gaps 0:

6 ATAACCCCAACCTATCGAAATATCGAATCTAGAATATATAAAACATTTCATTTTTAA 65

Db
417 ATATCTAATAGCCTACCAAGAAAGAGTAATTAAACCTAAAAAATTATTCGGCTAG 358

66 TGTTCCGTGTCGGTACTGCTACCAAAACAGAGGAGATAACAAGTGACACCGACCTTT 125

Db 357 ATTTCCTTTCTTTTCGTTTATCAATACACAAACTGAACCTACGACCAACATCACTT 298

QY 126 TAAGTCTTTTGGACTTACACCCTGAT 153

Db 297 T T C C T T T T T T T T C T T T T T C C T T T C T A A T 270

**0
0
0
1
3
0**

US-10-259-194A-119
Sequence 110 Analysis File 10050100A

Publication No. US20040010815A1
; GENERAL INFORMATION:

[illegible]

```

NAME/KEY misc feature
LOCATION: (4095)..(4369)
OTHER INFORMATION: "n" at positions 4099 to 4369 can be any base
FEATURE:
NAME/KEY: misc feature
LOCATION: (5502)..(5521)
OTHER INFORMATION: "n" at positions 5502 to 5521 can be any base
FEATURE:
NAME/KEY: misc feature
LOCATION: (10089)..(10620)
OTHER INFORMATION: "n" at positions 10089 to 10620 can be any base
FEATURE:
NAME/KEY: misc feature
LOCATION: (13273)..(13370)
OTHER INFORMATION: "n" at positions 13273 to 13370 can be any base
FEATURE:
NAME/KEY: misc feature
LOCATION: (20762)..(20781)
OTHER INFORMATION: "n" at positions 20762 to 20781 can be any base
FEATURE:
NAME/KEY: misc feature
LOCATION: (30751)..(30916)
OTHER INFORMATION: "n" at positions 30751 to 30916 can be any base
FEATURE:
NAME/KEY: misc feature
LOCATION: (46579)..(46772)
OTHER INFORMATION: "n" at positions 46759 to 46772 can be any base
FEATURE:
NAME/KEY: misc feature
LOCATION: (77382)..(77401)
OTHER INFORMATION: "n" at positions 77372 to 77401 can be any base
US-10-052-482-223

Query Match 4.7%; Score 35.2; DB 18; Length 79467;
Best Local Similarity 51.2%; Pred. No. 28;
Matches 82; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 428 TTTTGGCCAGCTTTGCTTTGCGAGTTTGGAGTGAAGTAAATCATTTTGTACGCTTTAG 487
Db 60135 TTTTATGCCCCGTGATGATAAACTAGCATGTGTGGCGAGGGCTGATGGCTGACAGAGCTGG 60076
QY 488 GCACTGTGCACTGTTTGTCTTGGCCGAACAACAGAGCGTTAAGCTGGGTAGTGGCGTCAAC 547
Db 60075 GAGCAGACACTGTCCGTTTCTTGGCCCAACCCAGACCATGCGCAGGAAATTGAAAGAAAGC 60016
QY 548 GTTTTGTGGCGATGTGGAGCGTTTGGCATGTGTCT 587
Db 60015 ATTTCCTACAGCTGTTGTGACTTGAGGCGCATGTGTCT 59976

RESULT 9
US-09-933-797-681/c
; Sequence 681, Application US/09933797
; Patent No. US20020155119A1
; GENERAL INFORMATION:
; APPLICANT: Robert A. Sikes et al.
; TITLE OF INVENTION: Isolation and use of fetal Urnental
; FILE REFERENCE: 9901-007-999
; CURRENT APPLICATION NUMBER: US/09/933,797
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: US/09/482,933
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: PCT/US99/10746
; PRIOR FILING DATE: 1999-05/14
; PRIOR APPLICATION NUMBER: 60/085,383
; NUMBER OF SEQ ID NOS: 811
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 681
; LENGTH: 214
; TYPE: DNA
; ORGANISM: Murine

```

US-09-933-797-681

Query Match 4.6%; Score 34.6; DB 9; Length 214;
Best Local Similarity 59.8%; Pred. No. 1.6;
Matches 56; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 7 TAAACCCCAACCTATCGAATAATCGAATCTAGAAATTAATAACATTCATTTTAAAT 66
DB 124 TAAGCAAAACCTGCTCTATGTAAGCTGAGAGACGATTAAGTCAAGTTTCTCAT 65

QY 67 GTTCGCTGCTGGGTACTGTCTACCAAAACAGAGAGA 103
DB 64 CTTCATGTGGGGACTAGGATCGAACTCAGAAACAA 28

RESULT 10
US-09-814-353-20855/c
Sequence 20855, Application US/09814353
Publication No. US20030165831A1
GENERAL INFORMATION:

APPLICANT: Lee, John
APPLICANT: Thompson, Pamela
APPLICANT: Lillie, James
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
FILE REFERENCE: MRI-006B

CURRENT APPLICATION NUMBER: US/09/814,353
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/191,031
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: US 60/207,124
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: US 60/211,940
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: US 60/216,820
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: US 60/220,661
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/257,672
NUMBER OF SEQ ID NOS: 22037
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20855
LENGTH: 3821
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 1, 64, 148, 3821
OTHER INFORMATION: n = A,T,C or G

US-09-814-353-20855

Query Match 4.6%; Score 34.6; DB 10; Length 3821;
Best Local Similarity 58.1%; Pred. No. 8;
Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 5 CATACCCCAACCTATCGAATAATCGAATCTAGAAATTAATAACATTCATTTTAA 64
DB 3799 CATTATCCCAACCTCTCATATCTCAACATGATTAAGAAATCATCATTTATGAT 3740

QY 65 ATGTCGCTGCTGGGTCTGTCTACCAAAACAGAGAGATTAACA 109
DB 3739 AGATTTAAGGTGTAATACAGGTTTCCAAAATGTGGCAGTGAGA 3695

RESULT 11
US-10-357-930-25623/c
Sequence 25623, Application US/10357930
Publication No. US20040255086A1
GENERAL INFORMATION:
APPLICANT: Schlegel, Robert
APPLICANT: Endege, Wilson

APPLICANT: Monahan, John
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
TITLE OF INVENTION: HUMAN PROSTATE CANCER
FILE REFERENCE: MRI-0078CN

CURRENT APPLICATION NUMBER: US/10/357,930
CURRENT FILING DATE: 2003-02-04
PRIOR APPLICATION NUMBER: 09/785,276
PRIOR FILING DATE: 2003-02-16
PRIOR APPLICATION NUMBER: 60/183,319
PRIOR FILING DATE: 2000-02-17
PRIOR APPLICATION NUMBER: 60/189,862
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/207,454
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: 60/211,314
PRIOR FILING DATE: 2000-06-09
PRIOR APPLICATION NUMBER: 60/219,007
PRIOR FILING DATE: 2000-07-18
PRIOR APPLICATION NUMBER: 60/255,281
PRIOR FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 62232
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 25623
LENGTH: 3825
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 1, 64, 148, 3825
OTHER INFORMATION: n = A,T,C or G

US-10-357-930-25623

Query Match 4.6%; Score 34.6; DB 20; Length 3825;
Best Local Similarity 58.1%; Pred. No. 8;
Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 5 CATACCCCAACCTATCGAATAATCGAATCTAGAAATTAATAACATTCATTTTAA 64
DB 3803 CATTATCCCAACCTCTCATATCTCAACATGATTAAGAAATCATCATTTATGAT 3744

QY 65 ATGTCGCTGCTGGGTACTGTCTACCAAAACAGAGAGATTAACA 109
DB 3743 AGATTTAAGGTGTAATACAGGTTTCCAAAATGTGGCAGTGAGA 3699

RESULT 12
US-10-437-963-75212/c
Sequence 75212, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 75212
LENGTH: 6630
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_75320C.1

US-10-437-963-75212

```
Query Match          4.6%; Score 34.4; DB 19; Length 6630;
Best Local Similarity 54.8%; Pred. No. 13;
Matches 66; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 528 GCTGGAGTAGTTGGCGTTCAGCGTTTGTCTGCGCATGATTTGGAGCGTTTGGCAATGTGCT 587
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3310 GCGGTGCGTTGGCGTGCAGCGAGGTGCGCGCGGTGAGTTGGCGCGCGCGCGGTGCG 3251

QY 588 GGGGCGTGGGGGGGCGCATCTGTTTCAGCATTTTTCGCCAGTATGTGCGCGAGTTAATA 647
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3250 GTGCGAGCGCGCGCGGAGGTGCGAGCGGTGCGCGCATGTGCGCGCATGTGCGCGGAAATC 3191

QY 648 TCGT 651
    |||
Db 3190 CCGT 3187

RESULT 13
US-10-260-238-2214/c
; Sequence 2214, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; PRIOR FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 2214
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: N_region
; LOCATION: (1919)..(1919)
; OTHER INFORMATION: n = any nucleotide
US-10-260-238-2214

Query Match          4.6%; Score 34.2; DB 17; Length 2000;
Best Local Similarity 56.8%; Pred. No. 7.6;
Matches 63; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 543 TCAGCGTTTCTGCGCATGATTTGGAGCGTTTGGCATGTGCTGGGCGCTGGCGGGGC 602
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1871 TGACCGAATTTGCGTGGGTTTGGGACTTTTGGGCTGGGCTGGGCTGTGATTTGGC 1812

QY 603 ATCGTTTACGCAATTTGTTGCGCATGTGTCGCCAGTATGTGCGCAATTAATATGTCG 653
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1811 GTCCCATGTGCAATGTGGAGGTTGGGCTGTGTCGCCAATGTCTGC 1761

RESULT 14
US-10-087-192-22/c
; Sequence 22, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.

Query Match          4.6%; Score 34.2; DB 13; Length 14163;
Best Local Similarity 47.1%; Pred. No. 83;
Matches 105; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 6 ATTAACCCAACTATCGAAATATGCAATCTAGATATATTAATAAATTGATTTTAA 65
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25440 ATCAGCGTCMAAACACTTAACCAACMAAGTTGAATGTAAACAGTACATATTTTAA 25381

QY 66 TGTTCGCGTGGGCTACTGTACCAAAACAGAGATTAACAGTACGACCGACCTTT 125
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25380 TATTAACCTTATTAACATGTTTATTAAGATTAATTCAGTCGACGTACCTCTACTAA 25321

QY 126 TAAATGCTTTTGAACCTTACACCGTATTAACGCTATGACGCCAGGACCGAACAATATTC 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25320 AAGTGCATGCTCTTGAACGCAATCTGCAATTTGTTCCAAAGAAACAATTCAGAAATC 25261

QY 186 TCCCGCTAGCTCTGCTAGCTGCGCATGATTCGTTGAAATGTC 228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25260 TGTGCAAACTTTAGTTAGATTAATGACCTTCATCAGATAC 25218

RESULT 15
US-10-424-599-84060/c
; Sequence 84060, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 84060
; LENGTH: 497
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_46923C.1
US-10-424-599-84060

Query Match          4.5%; Score 33.8; DB 18; Length 497;
Best Local Similarity 55.6%; Pred. No. 4.7;
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 257 TTTTGAATGATGATTAAGTGTGCGCGCATTTCAATTTCTACTGCGCATGATTAACCG 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 341 TGTATGCTGTGCTGTGCGGTGTGCTGTAAGCCCGCAACACTGCGCAAGATTTGCTCG 282

QY 317 GCAGCGTAAACCTTTTGAATTTGGCGGGCGCGCATATTTTCTGCGCTGGGCTGG 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 281 GGAGCTGTCAATTTTGTCTGCAAGTGGCGGATTTGTATTTAGATTTGCAATGCTGG 225
```

Search completed: September 10, 2005, 11:37:07
Job time : 629 secs

This Page Blank (uspto)

FEATURES	source
ORIGIN	
RESULT 2	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
FEATURES	
source	

Query Match	Similarity	46.8%;	Score 351;	DB 9;	Length 856;
Best Local	Similarity	100.0%;	Pred. No. 2.7e-95;		
Matches	351;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	400	AGACGACCTTCAGGCAAAACCAATCAGCTTTTGGGCGACGCTTTGCTTTGCAAGTTTGAA	459		
Db	1	AGACGACCTTCAGGCAAAACCAATCAGCTTTTGGGCGACGCTTTGCTTTGCAAGTTTGAA	60		
QY	460	CGTCAAAATATATTTTGTACGGGTGTACAGGCACTGTGACGTTTTGTTGCGCGAAACACA	519		
Db	61	CGTCAAAATATATTTTGTACGGGTGTACAGGCACTGTGACGTTTTGTTGCGCGAAACACA	120		
QY	520	GGCGTTAAGCTGGGATAGTTGAGCGTCAGCGTTTTGCTGGCATGATTTGGAGCGTTTGCAA	579		
Db	121	GGCGTTAAGCTGGGATAGTTGAGCGTCAGCGTTTTGCTGGCATGATTTGGAGCGTTTGCAA	180		
QY	580	TGTGTGCTGGCGCTGGCGGGGACATCTGTTTCAGCGATTTGTTTCGCCAGTAGTGGTCGCA	639		
Db	181	TGTGTGCTGGCGCTGGCGGGGACATCTGTTTCAGCGATTTGTTTCGCCAGTAGTGGTCGCA	240		
QY	640	GTTTAAATATGTGCTTGCCCTGTGTGCTGTGTCTATATGGCGGGTAGCGATTTTCTATTACG	699		
Db	241	GTTTAAATATGTGCTTGCCCTGTGTGCTGTGTCTATATGGCGGGTAGCGATTTTCTATTACG	300		
QY	700	AAAAAAAGCGAAGAGGTGCGCCCTTCGCGCTTAGTACTGTGCACTTAAAG	750		
Db	301	AAAAAAAGCGAAGAGGTGCGCCCTTCGCGCTTAGTACTGTGCACTTAAAG	351		

LOCUS	CL688782	791 bp	DNA	linear	GSS 09-JUL-2004
DEFINITION	PRIO14b_G04_2 - PRIO14b BR (791) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.				
ACCESSION	CL688782				
VERSION	CL688782.1	GI:50197860			
KEYWORDS	GSS.				
SOURCE	Pristionchus pacificus				
ORGANISM	Pristionchus pacificus				
	Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.				
REFERENCE	1 (bases 1 to 791)				
AUTHORS	Strinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.U.				
TITLE	Appads: an Acedb database for the nematode satellite organism Pristionchus pacificus				
JOURNAL	Pucleic Acids Res. 32 (1), D421-D422 (2004)				
COMMENT	Contact: Sommer RJ Evolutionary Biology Max-Planck-Institute for Developmental Biology Spemannstr. 37-39, Tuebingen D-72076, Germany Tel.: 00497071601371 Fax: 00497071601498 Email: ralf.sommer@uebingen.mpg.de This library was generated at Caltech, Pasadena, USA and end sequenced at Vancouver, Canada. Seq primer: T7 Class: fosmid ends.				
FEATURES	location/Qualifiers				
Source	1..791				
	/organism="Pristionchus pacificus"				
	/mol_type="genomic DNA"				
	/strain="California"				
	/db_xref="taxon:54126"				
	/clone_lib="Mixed stage fosmid library of P. pacificus var. California"				
	/note="Vector: pBpifos-5 Fosmid vector"				
ORIGIN					
Query Match	44.2%	Score 331.4;	DB 9;	Length 791;	
Best Local Similarity	99.7%	Pred. No. 2,5e-89;			
Matches 332; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0	

QY 418 ACCAATCAGCTTTGGGCGACGTTTGGCTTGCAATTTGTGAAGCTCAAAATCATTTTGTGA 477
 Db 3 ACRATACGCTTTTGGGCGACGTTTGGCTTGCAATTTGTGAAGCTCAAAATCATTTTGTGA 62
 QY 478 CGGTGTTACGGCACTGTCAGAGTTTGTTCGCCGCAACAACAGCGCTTAAAGCTGGAGT 537
 Db 63 CGGTGTTACGGCACTGTCAGAGTTTGTTCGCCGCAACAACAGCGCTTAAAGCTGGAGT 122
 QY 538 TGGCGTACAGGCTTTGCTGCGAGATGATTTGGGACGTTTGGCAATGTGCTGGCGCTGGC 597
 Db 123 TGGCGTACAGGCTTTGCTGCGAGATGATTTGGGACGTTTGGCAATGTGCTGGCGCTGGC 182
 QY 598 GGGGCACTGTTTTCAGGAGATTTGTCGCGAGATGATGCGCAGTTAAATATCGTGTGC 657
 Db 183 GGGGCACTGTTTTCAGGAGATTTGTCGCGAGATGATGCGCAGTTAAATATCGTGTGC 242
 QY 658 CTTGTGCTGCTATTTGCGCGGTACGCAATTTTCTATTAACGAAAAAAGCGAAGAGT 717
 Db 243 CTTGTGCTGCTATTTGCGCGGTACGCAATTTTCTATTAACGAAAAAAGCGAAGAGT 302
 QY 718 CGCCTCTTCCGCTTAACTTGTCTTAAAG 750
 Db 303 CGCCTCTTCCGCTTAACTTGTCTTAAAG 335

RESULT 4
 CL688905 736 bp DNA linear GSS 09-JUL-2004
 LOCUS PR1014d.C02.2 - PR1014d.BR (736) Mixed stage fosmid library of P.
 DEFINITION Pacificus var. California Pristionchus pacificus genomic, genomic
 survey sequence.
 ACCESSION CL688905
 VERSION CL688905.1 GI:50198003
 KEYWORDS GSS.
 SOURCE Pristionchus pacificus
 ORGANISM Pristionchus pacificus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
 Neodiplogasteridae; Pristionchus.
 1 (bases 1 to 736)
 AUTHORS Strinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
 TITLE Appabdi: an Acedb database for the nematode satellite organism
 Pristionchus pacificus
 JOURNAL Nucleic Acids Res. 32 (1), D421-D422 (2004)
 COMMENT Contact: Sommer RJ
 Evolutionary Biology
 Max-Planck-Institute for Developmental Biology
 Spemannstr. 37-39, Tuebingen D-72076, Germany
 Tel: 00497071601371
 Fax: 00497071601498
 Email: ralf.sommer@uebingen.mpg.de
 This library was generated at Caltech, Pasadena, USA and end
 sequenced at Vancouver, Canada.
 Seq primer: T7
 Class: fosmid ends.
 Location/Qualifiers
 1..736
 /organism="Pristionchus pacificus"
 /mol_type="genomic DNA"
 /strain="California"
 /db_xref="taxon:54126"
 /clone_lib="Mixed stage fosmid library of P. pacificus
 var. California"
 /note="Vector: pEplfos-5 Fosmid vector"

ORIGIN

Query Match 42.9%; Score 321.6; DB 9; Length 736;
 Best Local Similarity 98.8%; Pred. No. 2,4e-86;
 Matches 32; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 422 ATCAGCTTTTGGGCGACGTTTGGCTTGCAATTTGTGAAGCTCAAAATCATTTTGTGAAGT 481
 Db 6 ATCAGCTTTTGGGCGACGTTTGGCTTGCAATTTGTGAAGCTCAAAATCATTTTGTGAAGT 65

QY 482 GTTACGGCACTGTCAGAGTTTGTTCGCCGCAACAACAGCGCTTAAAGCTGGAGTGGC 541
 Db 66 GTTACGGCACTGTCAGAGTTTGTTCGCCGCAACAACAGCGCTTAAAGCTGGAGTGGC 125
 QY 542 GTTACGGCTTTTGGCTGCGAGATGATTTGGGACGTTTGGCAATGTGCTGGCGCTGGC 601
 Db 126 GTTACGGCTTTTGGCTGCGAGATGATTTGGGACGTTTGGCAATGTGCTGGCGCTGGC 185
 QY 602 CATCTGTTTACGAGATTTTGGCGCAGTATGTCGCGCAATTAATATCGTGTGGCTGG 661
 Db 186 CATCTGTTTACGAGATTTTGGCGCAGTATGTCGCGCAATTAATATCGTGTGGCTGG 245
 QY 662 TTGCTGCTATTTGCGCGGTACGCAATTTTCTATTAACGAAAAAAGCGAAGAGTGGC 721
 Db 246 TTGCTGCTATTTGCGCGGTACGCAATTTTCTATTAACGAAAAAAGCGAAGAGTGGC 305
 QY 722 CTCCTCCGCTTAACTTGTCTTAAAG 749
 Db 306 CTCCTCCGCTTAACTTGTCTTAAAG 333

RESULT 5
 CA374430 118 bp mRNA linear EST 06-NOV-2002
 LOCUS 648800 NCCCA 1RT Oncorhynchus mykiss cDNA clone 1RT18B19_B.10 5',
 DEFINITION mRNA sequence.
 ACCESSION CA374430
 VERSION CA374430.1 GI:24690829
 KEYWORDS EST.
 SOURCE Oncorhynchus mykiss (rainbow trout)
 ORGANISM Oncorhynchus mykiss
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 1 (bases 1 to 118)
 AUTHORS Rexroad,C.E. 3rd, Lee,Y., Keeler,J.W., Karamycheva,S., Brown,G.,
 Koop,B., Gahr,S.A., Palti,Y. and Quackenbush,J.
 TITLE Sequence analysis of a rainbow trout cDNA library and creation of a
 gene index
 JOURNAL Cytogenet. Genome Res. 102 (1-4), 347-354 (2003)
 COMMENT Contact: Rexroad CE
 USDA, ARS, National Center for Cool and Cold Water Aquaculture
 11876 Lactown Road, Kearneyville, WV 25430, USA
 Tel: 304 724 8340 x2129
 Fax: 304 725 0351
 Email: crexroad@nccwa.ars.usda.gov
 Single pass sequencing. Bases called with phred v0.020425.c and
 trimmed with the aid of the trim_alt option. Vector identified by
 cross match v0.990329.
 Seq primer: AGCGATTAACATTTTCACACAGA.
 Location/Qualifiers
 1..118
 /organism="Oncorhynchus mykiss"
 /mol_type="mRNA"
 /db_xref="taxon:8022"
 /clone_lib="1RT18B19_B.10"
 /cissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="NCCCA 1RT"
 /note="Vector: PCMV SPORT6; Site 1: NotI; Site 2: SalI;
 library made from pooled tissue from brain, gill, liver,
 spleen, muscle, and kidney."

ORIGIN

Query Match 15.1%; Score 113.2; DB 6; Length 118;
 Best Local Similarity 97.5%; Pred. No. 4.5e-23;
 Matches 115; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 336 GTTGGCGGGGCGGCGCATATTTGTGCTGCGCTGGCGGAAATCGCCACGACCCCAACA 395
 Db 118 GTTGGCGGGGCGGCGCATATTTGTGCTGCGCTGGCGGAAATCGCCACGACCCCAACA 59
 QY 396 AGGAAGACGAGCTTCAGGCAACAAATCAGCTTTTGGGCGAGCTTTGCTTGGCAATT 453

Db 58 AGAAGACGGGCTTCAGGCAAAACCATCAGCTTTTGGCCAGCTTTGCTTGAGATT 1

RESULT 6
BH447440/c 819 bp DNA linear GSS 12-DEC-2001
BOH0849TF BOHQ Brassica oleracea genomic clone BOH0849, genomic
survey sequence.

ACCESSION BH447440
VERSION BH447440.1 GI:17633154
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 819)
Town, C.D., Van Aken, S., Utecherback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_GSSs: BOH0849TF
COMMENT Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES
source location/Qualifiers
1..819
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOH0849"
/clone_1lb="BOHQ"
/note="Vector: PHOSI, Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOSI using BstXI linkers"

ORIGIN
Query Match 7.4%; Score 55.2; DB 8; Length 819;
Best Local Similarity 44.8%; Pred. No. 3.4e-05;
Matches 256; Conservative 0; Mismatches 313; Indels 3; Gaps 1;

QY 122 CTTTAAAGTCTTTTGGACTTACACCTGATTAACCGCTATGACGCGACGACAT 181
DB 734 CCTATATTTCTTTTGTATTTGTTTATTAAGTAGCTTCACACCGGACCTTAACAT 675
QY 182 ATTTCGCGCTTAGCTCTGCTACGTCGATGAGATTTGCTCAAGTACCGGCGTGGCA 241
DB 674 TTTTAAAGGATGACATACGCTTAACATGCTTTTAAAGATGATGCAATTTTGCTTT 615
QY 242 GGAATGAGTCGGGATTTTGAATGTATGTATCTGTGCGGGCATTTTCACTG 301
DB 614 GGAAGAGTTTCGGGTTTTCATCTCACTTCCTTAATGAGCTTTTAATTTTCTTA 555
QY 302 GCAGTATGTAACCCGCGACCGGTACACCTTTGAGTTGGGGCGGCGCATATATTGTC 361
DB 554 ATTAAGATTACCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATGCTC 495
QY 362 TGGCTGGGGTGAATAATGCGACCAACGCAAGAAAGCGAATTCAGGCA---AAA 418
DB 494 TATTAGCTTTTAAATACTTAACAAGTAAACCTAGCACAGATCCGATGAAAAATATAT 435
QY 419 CCAATGAGCTTTTGGGCGCAGCTTTGCTTGGAGTTTGAAGTCAAAATCATTTTGTAC 478
DB 434 AAAAATTATTATACAGTAGAGATTCCTTCAAGTTTATTAACCTTAAGGATATCTATTC 375
QY 479 GGTGTACGAGACTGTGACGTTTGTCTGCGCAACACAGGCGTTAAGCTGGGTAGTT 538

Db 374 GGACTAACCGTAGTATCACTTTATTTCTCCCTTATTAACAATTCATATTCAGATTATCTA 315

QY 539 GGGCTGACGCTTTTGTGCGCATGATGATGGAGCTTTGGCAATGTGCTGGGCGTGGG 598

Db 314 CTTTTCATTTATTTCTCGGTGAGTTAATGATGATGATGATGATGATGATGATGAT 255

QY 599 GGGCATCTGTTTTCAGCATTTGTTTCGCCAGTATGTCGCCAGTTAAATATCGCTTGGC 658

Db 254 GGTTCATTTTCAAAACATATTATTAACATACATTAATTCATTTAATCATATTATGGCC 195

QY 659 CTGTGCTGTGCTATTGTCGGCGGTAGCATTTT 690

Db 194 GTTTTATGATTGTGATGCAATTCATTTGT 163

RESULT 7
BH374397 420 bp DNA linear GSS 10-DEC-2001
LOCUS BH374397
DEFINITION AG-ND-180D20.TF ND-TAM Anopheles gambiae genomic clone
AG-ND-180D20, genomic survey sequence.

ACCESSION BH374397
VERSION BH374397.1 GI:17320539
KEYWORDS GSS.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
1 (bases 1 to 420)
Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J.,
Ren, C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H., B.,
Gardner, M.J. and Collins, F.H.
Construction of a BAC library and generation of BAC end
sequence-tagged connectors for genome sequencing of the African
malaria mosquito Anopheles gambiae
Mol. Genet. Genomics 268 (6), 720-728 (2003)
22542063
12655398
Other_GSSs: AG-ND-180D20.TF
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 Rev
Class: BAC ends.

FEATURES
source location/Qualifiers
1..420
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-180D20"
/clone_1lb="ND-TAM"
/note="Vector: pECBAC1, Site_1: HindIII"

ORIGIN
Query Match 6.7%; Score 50; DB 8; Length 420;
Best Local Similarity 49.4%; Pred. No. 0.0011;
Matches 159; Conservative 0; Mismatches 160; Indels 3; Gaps 1;

QY 147 CCTGATTACCGCTATGACCGCAGACCGAACAATATTCTGCGCTTAGCTGTGCTACGT 206

Db 93 CTTTGTGACGCTGCTACCGCCGCGCCCAACATGATGCTACGCGCTTCGCGGGCCCA 152

Qy 207 CGATGAGATTTTCGTAAGTACCGCGCTGCTGCGAGGATGATGCTGGATTTTGTATG 266

Db 153 CTTTGGCTACCGCGGATTTGTGCGCATATGCTGGCATCTGCTGGGGGTGTGTATCA 212

Qy 267 TGAATGTAATGCTGTGCGGCGATTTTCACTGCGAGTATGATACCGCGACGGGTAC 326

Db 213 TGGTGTGCTGATCGCGGCGCGGCTGCGCCAGCTGTTCGAGGCCGAACCGGCACTTCA 272

Qy 327 ACCTTTGTAGTTGGGCGGCGGCGCATATATTTGCTGGCTGCGCTGGAAAATGCGCACCA 386

Db 273 CCTCTGCTCAATATGACAGGTGCGCGCTGATCTGCTGCTGCGCTGGAAATGCGCGCTG 332

Qy 387 GCCCAACAAAGAGACGAGCTTCAGGCAAAA--CCAAATCAGCTTTTGGCGCAGCTTTG 443

Db 333 CCGGCAAGCGTGAATCAGGAGACAGTCAAGCAACGCTCGCTCAGCTTCTGGGAGCGCGAG 392

Qy 444 CTTTGCAGTTTGTGACAGTCA 465

Db 393 CTTTCCAGTGGGTCAACCCCA 414

RESULT 8
CL666889/c 862 bp DNA linear GSS 09-JUL-2004

LOCUS PR10153C.G02 - PR10153C.B21 (862) Mixed stage fosmid library of P. pacificus var. California Pristionchus pacificus genomic, genomic survey sequence.

ACCESSION CL666889.1 GI:50160241

VERSION GSS.

KEYWORDS Pristionchus pacificus

SOURCE Pristionchus pacificus

ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Diphlogasterida; Neodiphlogasteridae; Pristionchus.

REFERENCE 1 (bases 1 to 862)
Strinivasan,V., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J. AppABD: an Acedb database for the nematode satellite organism Pristionchus pacificus

JOURNAL Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.

FEATURES
source location/Qualifiers
1..862
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus var. California"
/note="Vector: pBplfos-5 Fosmid vector"

ORIGIN
Query Match 6.6%; Score 49.4; DB 9; Length 862;
Best Local Similarity 48.7%; Pred. No. 0.002;
Matches 133; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

Qy 111 TGACACGACCCCTTTAAGCTTTTGGACTTACACCCGATACCGGCTATGACGCCAG 170

Db 460 TGTGACACGACCTTTATCGGCCCTTGTGCTGTGTGCTTCCTTTCGATCAGCGCCG 401

Qy 171 GACGAAACATATTTCTGCCCTTAGCTGTGCTACGTGACATGATTTGTCGAAAGTACC 230

Db 400 GTCCGAATACACCATGACTACTCGCTCGGCGCTCACTTCGCTTCGCTGCATGC 341

Qy 231 GCGTGTGCGAGGATGATGCTGGATTTTGTATGATGATGATGATGATGATGATGATG 290

Db 340 CCATGAGCTGGGATCAGATCGGCTTCACTGCTGGTATGATGATGATGATGATGATG 281

Qy 291 CATTTTCACTGCGAGTATGATCCGCGACCGGTATACCTTTGATGATGATGATGATG 350

Db 280 TGGCGAGAGTGTTCAGATATTCCTTGGGCTTACACGCTGTGCGCTATGATGATG 221

Qy 351 CATATATTTGCTGCTGCGGCTGGAAAATGCGCAC 385

Db 220 GATACCTGCTGTACTGCGCTGGAATGCGCAC 186

RESULT 9
BH379453 450 bp DNA linear GSS 10-DEC-2001

LOCUS AG-ND-163M3.TR ND-TAM Anopheles gambiae genomic clone AG-ND-163M3, genomic survey sequence.

ACCESSION BH379453

VERSION BH379453.1 GI:17325595

KEYWORDS GSS.

SOURCE Anopheles gambiae (African malaria mosquito)

ORGANISM Anopheles gambiae

REFERENCE 1 (bases 1 to 450)
Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J., Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B., Gardner,M.V. and Collins,F.H. Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito Anopheles gambiae

JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)

COMMENT MEDLINE
PUBMED 12655398
Other GSSs: AG-ND-163M3.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PBST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 Rev
Class: BAC ends.

FEATURES
source location/Qualifiers
1..450
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PBST"
/db_xref="taxon:7165"
/clone_lib="AG-ND-163M3"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1, Site_1: HindIII"

ORIGIN
Query Match 5.8%; Score 43.4; DB 8; Length 450;
Best Local Similarity 47.9%; Pred. No. 0.12;
Matches 125; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 147 CCCGATTACCGCTATGACCGCAGACCGAACAATATTCTGCGCTTACGCTGTCTACGT 206
DB 68 CTTTGTGACGTCCTGTATGACCGCGCCCAACATGATGCTGACCGGCTTCGGGGGCCA 127
QY 207 CGCATGATTTTCGTCAAAAGTACCCGCTGCTGCGCAGGATGATGCTGGGATTTTGTGATTTG 266
DB 128 CTTTCGGCTACCGCGCAGCGGTGCGGATATGCTGGGCATCTGCTGGGGGTGGTATCA 187
QY 267 TGATGTATCTGTGTGCGGGCATTTTCATTTTCATGCGAGTGTGACCCGGCAGCGGTAC 326
DB 188 TGTGTGCTGTATGCGCGCGGCTGCGCAGCTGTTCAGAGCCGAACCGCGCATCTTACA 247
QY 327 ACCTTTTGTGATTTGGCGCGCGCATATATGCTGCTGCGCTGCGGAAATCGCCACCA 386
DB 248 CCCGCTCATTTATGTTGTGCGCGCTGTCTGCTGCTGCGCTGCGAAGATTGCCGTG 307
QY 387 GCGCCACAAAGAAAGACGAC 407
DB 308 CCGCAGCTGTGATCAGGAC 328

RESULT 10
CL139330/c 938 bp DNA linear GSS 05-JAN-2004
LOCUS ISB1-112B4.T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-112B4,
DEFINITION genomic survey sequence.
ACCESSION CL139330
VERSION CL139330.1 GI:40632965
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 938)
Kremetzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
A physical map of the xenopus tropicalis genome
TITLE Unpublished (2003)
JOURNAL Contact: Richard K Wilson
COMMENT Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: T7 TAATACGACTCATATAGCG
Class: BAC ends
High quality sequence start: 2
High quality sequence stop: 794.
Location/Qualifiers
1. .938
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-112B4"
/clone_11b="ISB1"
/note="Vector: pBelOBAC11; ISB-1 Xenopus tropicalis BAC
Library Segment 1"

ORIGIN
Query Match 5.2%; Score 39.2; DB 9; Length 938;
Best Local Similarity 56.1%; Pred. No. 2.7;
Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 35 TCTTGATATATAAACATCTTTTAAATGTTCCGTGCGGTACTGTCTACCAAA 94
DB 810 TCAGGAATACCTAGGCTTCTGTATGTTGATGATGCCAATAATAGCTTATATAA 751
QY 95 CAGAGAGATAAAGTACACCGACCTTTAAGTGTCTTTTGAAGTACACCTGATT 154
DB 750 GCAATGTTTACACTGGGCAAGACCCCAAGAGAGCTTTTAAACTACAGACTGANTG 691
QY 155 ACCGCTATGACG 166
||| |||||

DB 690 ACCAGCTTGACG 679

RESULT 11
AG240288/c 474 bp DNA linear GSS 19-JUL-2003
LOCUS Lotus corniculatus var. japonicus DNA, clone:Lj106d06_not, genomic
DEFINITION survey sequence.
ACCESSION AG240288
VERSION AG240288.1 GI:26551125
KEYWORDS GSS.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM
Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteeae;
Lotus.
REFERENCE 1
Sato, S., Nakamura, Y. and Tabata, S.
Lotus japonicus TAC End sequences
TITLE Published Only in Database (2002)
JOURNAL 2 (bases 1 to 474)
REFERENCE Sato, S.
TITLE Direct Submission
JOURNAL Submitted (20-NOV-2002) Shusei Sato, Kazusa DNA Research Institute,
The First Laboratory for Plant Gene Research; 2-6-7
Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
E-mail: ssato@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/,
Tel: 81-438-52-3935 (ex.2336), Fax: 81-438-52-3934
Location/Qualifiers
1. .474
/organism="Lotus corniculatus var. japonicus"
/mol_type="genomic DNA"
/strain="Miyakojima MG-20"
/variety="japonicus"
/db_xref="taxon:34305"
/clone="Lj106d06 not"
/clone_11b="genomic TAC library"
/note="VECTOR: pYL1TAC7-synonym: Lotus japonicus"

ORIGIN
Query Match 5.1%; Score 38.2; DB 9; Length 474;
Best Local Similarity 51.5%; Pred. No. 4.5;
Matches 88; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 579 ATGTGCTCGGGCGGCTGCGGCGCATCTTTAGGAGATTGTTCCGCAATGTCGCC 638
DB 307 ATGGGTGTGGGATCTGCGCGGTCTTTGTCTACCTAGTTCACACTGTCTATTG 248
QY 639 AGTTAAATATCGTCTTGCCCTGTGCTGATTTGCGGCTACGCAATTTCTATTAC 698
DB 247 TTTTCAGGGCAGTCTGTCTGTCTGATGATGATTTGGGAACCAATTTCAATAG 188
QY 699 GAAAAAAGCGAAGAGTCCCTTTCCGCTTACTGACTTCTTAA 749
DB 187 CTTCTGCGGATTTGTCACCTTCCCTTTCACATAAACAATCTGCTTAA 137
||| |||||

RESULT 12
CNS00BEC 1101 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR29104 of RPCL1-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL069085
VERSION AL069085.1 GI:4949228
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)

McClister, V., Clinton, S., Canapelli, B., Page, D., Martin, J.,
Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Meising, B.,
Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C.,
Tasgersishvili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C.,
Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T.,

McCarter, J., Clifton, S., Chiappelli, B., Page, D., Martin, J.,
Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B.,
Bowers, Y., Gibbons, M., Ritter, B., Bennett, J., Franklin, C.,
Tengstedt, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C.,
Tsagaraki, A., Allen, M., Pearson, B., Swaller, T.,
Tsagaraki, A., Sceptoe, M., Allen, M., Pearson, B., Swaller, T.,

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2005, 01:36:31 ; Search time 42 Seconds

(without alignments)
346.585 Million cell updates/sec

Title: US-10-620-487-2

Perfect score: 996

Sequence: 1 VTPTLSAFWTYTLITAMTP.....RQNLIVALLLVYCAVRIFY 195

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents AA:*

1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*

2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*

3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*

4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*

5: /cgn2_6/prodata/1/1aa/6C.COMB.pep:*

6: /cgn2_6/prodata/1/1aa/backfile1.pep:*

7: /cgn2_6/prodata/1/1aa/backfile1.pep:*

8: /cgn2_6/prodata/1/1aa/backfile1.pep:*

9: /cgn2_6/prodata/1/1aa/backfile1.pep:*

10: /cgn2_6/prodata/1/1aa/backfile1.pep:*

11: /cgn2_6/prodata/1/1aa/backfile1.pep:*

12: /cgn2_6/prodata/1/1aa/backfile1.pep:*

13: /cgn2_6/prodata/1/1aa/backfile1.pep:*

14: /cgn2_6/prodata/1/1aa/backfile1.pep:*

15: /cgn2_6/prodata/1/1aa/backfile1.pep:*

16: /cgn2_6/prodata/1/1aa/backfile1.pep:*

17: /cgn2_6/prodata/1/1aa/backfile1.pep:*

18: /cgn2_6/prodata/1/1aa/backfile1.pep:*

19: /cgn2_6/prodata/1/1aa/backfile1.pep:*

20: /cgn2_6/prodata/1/1aa/backfile1.pep:*

21: /cgn2_6/prodata/1/1aa/backfile1.pep:*

22: /cgn2_6/prodata/1/1aa/backfile1.pep:*

23: /cgn2_6/prodata/1/1aa/backfile1.pep:*

24: /cgn2_6/prodata/1/1aa/backfile1.pep:*

25: /cgn2_6/prodata/1/1aa/backfile1.pep:*

26: /cgn2_6/prodata/1/1aa/backfile1.pep:*

27: /cgn2_6/prodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	684	68.7	205	US-09-489-039A-12661	Sequence 12661, A
2	272.5	27.4	249	US-09-252-991A-25717	Sequence 25717, A
3	250.5	25.2	211	US-09-489-039A-10442	Sequence 10442, A
4	232.5	23.3	212	US-09-543-681A-4767	Sequence 4767, Ap
5	228.5	22.9	211	US-09-543-681A-6708	Sequence 6708, Ap
6	191.5	19.2	238	US-09-328-352-7122	Sequence 7122, Ap
7	178	17.9	241	US-09-328-352-7159	Sequence 7159, Ap
8	173	17.4	271	US-09-252-991A-32455	Sequence 32455, A
9	163.5	16.4	235	US-09-328-352-8249	Sequence 8249, Ap
10	162	16.3	214	US-09-489-039A-8066	Sequence 8066, Ap
11	158.5	15.9	206	US-09-252-991A-18746	Sequence 18746, A
12	155.5	15.6	214	US-09-328-352-6025	Sequence 6025, Ap
13	155.5	15.6	237	US-09-252-991A-27175	Sequence 27175, A
14	154	15.4	210	US-09-328-352-6808	Sequence 6808, Ap
15	153.5	15.4	211	US-09-252-991A-20747	Sequence 20747, A
16	153	15.4	214	US-09-489-039A-9793	Sequence 9793, Ap
17	151.5	15.2	211	US-09-902-540-9913	Sequence 9913, Ap
18	149.5	15.0	250	US-09-543-681A-4487	Sequence 4487, Ap
19	148	14.9	211	US-09-489-039A-9608	Sequence 9608, Ap
20	146.5	14.7	220	US-09-328-352-4903	Sequence 4903, Ap
21	146	14.7	222	US-09-328-352-6471	Sequence 6471, Ap
22	138.5	13.9	219	US-09-489-039A-7731	Sequence 7731, Ap
23	132	13.3	208	US-09-328-352-5111	Sequence 5111, Ap
24	128.5	12.9	228	US-09-543-681A-4854	Sequence 4854, Ap
25	128.5	12.9	240	US-09-328-352-7537	Sequence 7537, Ap
26	128	12.9	220	US-09-543-681A-4545	Sequence 4545, Ap
27	127	12.8	200	US-09-328-352-4773	Sequence 4773, Ap

28	126	12.7	205	3	US-09-396-357-2	Sequence 2, Appl1
29	125.5	12.6	214	4	US-09-328-352-6695	Sequence 6695, Ap
30	123.5	12.4	228	4	US-09-252-991A-23837	Sequence 23837, A
31	123.5	12.4	277	4	US-09-252-991A-23018	Sequence 23018, A
32	122.5	12.3	201	4	US-09-328-352-6409	Sequence 6409, Ap
33	115.5	11.6	260	4	US-09-252-991A-25992	Sequence 25992, A
34	112.5	11.3	249	4	US-09-328-352-5813	Sequence 5813, Ap
35	112	11.2	217	4	US-09-489-039A-8076	Sequence 8076, Ap
36	111.5	11.2	222	4	US-09-252-991A-26644	Sequence 26644, A
37	109.5	11.0	226	4	US-09-328-352-4784	Sequence 18666, A
38	107.5	10.8	206	4	US-09-328-352-8225	Sequence 4784, Ap
39	105.5	10.6	209	4	US-09-328-352-8225	Sequence 8225, Ap
40	104.5	10.5	210	4	US-09-328-352-4775	Sequence 4775, Ap
41	103.5	10.4	226	4	US-09-252-991A-31610	Sequence 31610, A
42	102.5	10.3	221	4	US-09-328-352-6659	Sequence 6659, Ap
43	99	9.9	213	4	US-09-489-039A-11845	Sequence 11845, A
44	98.5	9.9	138	4	US-09-711-164-445	Sequence 445, Ap
45	98.5	9.9	269	4	US-09-489-039A-7320	Sequence 7320, Ap

ALIGNMENTS

```
RESULT 1
US-09-489-039A-12661
Sequence 12661, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489, 039A
PRIORITY FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117, 747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 12661
LENGTH: 205
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12661

Query Match      68.7%; Score 684; DB 4; Length 205;
Best Local Similarity 68.4%; Pred. No. 3.4e-68;
Matches 132; Conservative 21; Mismatches 40; Indels 0; Gaps 0;

QY      1 VTPTLSAFWTYTLITAMTPGPNNIIALSATSHGFROSTRVLGMSIGFLIVMLCAGI 60
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      10 VTPTLSAFWTYTLITAMTPGPNNIIALSATSHGFROSTRVLGMSIGFLIVMLCAGI 69
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      61 SFLAVIDPAVAHLISWAGAAVYVWLAKIATSPYKEDGIQAKPISFASPALQPVNVI 120
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      70 TFLVLEDSRFTLVIGMGAAYILMLAMQIAKSPATGTPSVFPMFMSLIGQFVNKI 129
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      121 ILTGWALSPFVPOQALSWVGVSVLAMICTFPGVNCALGHI FORFROYGRONT 180
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      130 ILTGWALSPFVPOQALSWVGVSVLAMICTFPGVNCALGHI FORFROYGRONT 189
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      181 VLLALLVYCAVRI 193
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      190 MLLALLVYCAVRI 202
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 2
US-09-252-991A-25717
Sequence 25717, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
```

```

: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 25717
: LENGTH: 249
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-25717

```

Query Match	27.4%;	Score 272.5;	DB 4;	Length 249;
Best Local Similarity	37.0%;	Pred. No. 3.1e-22;		
Matches 71;	Conservative 39;	Mismatches 67;	Indels 15;	Gaps 8;

QY 5 LLSAEFTYLLITAMTPGPNNNILASSATSGFROSTRVLAMSGFLIWLKAGISFSL 64
 Db 51 LLLAASLSEFVTSVTPGPNNTMLLAAGVAFGVSIPHILGISCFPIWMA--AVGFL 107

Cy 65 AVI--DPAAVHLSMAGAAIYIWLAMKATIS-PRKEEDG-QAQRISFMSFAIQEVNPK 119
 :
 Db 108 GIIVFEAYPLVTILIRYGAAILLIYIAKKATISGPAGNDPEGRGRPLSYMGAAAFQWVNPK 167

[illegible]

Oy	176	ROLNIVLALLLV	187
	↓	: : : : : ↓	
Db	225	RVENVMALLLV	236

RESULT 3
US-09-489-039A-10442
S. Serenice 10442 Application IIS/09489039A

; PATENT NO. 6610836
 ;
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION. NITRIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBS

```

1  TITLE OF INVENTION:  ENECROMIOL FOR DIAGNOSTICS AND THERAPEUTICS
2
3  FILE REFERENCE:  2709.2004001
4
5  CURRENT APPLICATION NUMBER:  US/09/489,039A
6
7  CURRENT FILING DATE:  2000-01-27

```

[illegible]

```

      /
      / TYPE: PRT
      /
      / ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10442

```

Query Match	25.2%	Score 250.5;	DB 4;	Length 211;
Best Local Similarity	32.4%	Pred. No. 7e-20;		
Matches 61; Conservative	42;	Mismatches 78;	Indels 7;	Gaps 5

[illegible]

QY 65 AVIDAAVHL-LSMGAAYIVLANKIATS--PTKEDGLAKPISEWASFALQFNVIKI 121
:::|||||:::|||||:
Db 76 VERNDDVVYLIRICGATLYLLYLAANIATAFLPMKSISESSARPFSEWEAGQWNPKAM 135

Db

122	LYGVATSTFVLPFOALSWVGVSYLLAMGTFCNVCMALAGHLEFRLFR--QYGRQLN	179
136	IMATGATITTY-LPQGENVRSVIFLSLLFASINAPSVSIWTFGATLRHMLNDIKYLRIFN	194

QY	180	IVLALTLV	187
	:	:	:
Db	195	VCMMAMILL	202

RESULT 4
US-09-543-681A--4767
; Sequence 4767, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRAB
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543, 681A

PRIOR APPLICATION NUMBER: US 60/128, 706
PRIOR FILING DATE: 1993-04-09
NUMBER OF SEQ ID NOS: 8344

```

;
; LENGTH: 212
;
; TYPE: PRT
;
; ORGANISM: Proteus mirabilis
;

```

Query Match	Score	DB 4	Length	212;
Best Local Similarity	30.23;	Pred. No. 7.1e-18;		
Best Similarity	45.77;	Midnight	7.	Cans
		Totals	6.	

QY 12 YLTITAMTPGNNILATSSATSHGFROSTRVLAGMSGLFLIWLMLCAGISF-SLAVIDPA 70
:::|||||:::|:::
26 PRRFTCTCPGRDNIWTLASCINECIKPTMDHATGVSGIF-FWTLLVAGTGGALIRSPPI A3
:::|||||:::|:::

Qy 71 AHLLSNAGAAYYIVMLAMKXIATSPKE--DGLQAKPISFWASPALQFVNKKILYGVATAL 128
.: : : : : : : : : : : : : : : :
Db 84 TNYIKVIGAYITWIAWKTAISSHVEONSSNKGSPITLLEALFOVINPSMMMAISGI 143
.: : : : : : : : : : : : : : : :

Qy 129 STFFVLPTQALSMVVGVSLLIAMIGTFEGNVCMALAGH LFFQRLFR QYGRQQLNTVALLL 186

Db 144 TYTSTPOXPYPSMLL VAIIFLLINFPQCAIMTFEGHSLRERIKNPKILKLFNFINGLL 202

QY	187	VYCAVRIFY	195
		: : :	
Db	203	ALSAISVLF	211

RESULT 5
US-09-543-681A-6708

```

; Patent No. 6605709
;
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
;

```

```

; TITLE OF INVENTION:  DIAGNOSTICS AND THERAPEUTICS
;
; FILE REFERENCE:  2709.1002-001
;
; CURRENT APPLICATION NUMBER:  US/09/543,681A

```

```

; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1993-04-09
; NUMBER OF SEQ ID NOS: 8344

```

```
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Proteus mirabilis
```

Query Match	22.9%	Score 228.5;	DB 4;	Length 211;
Best Local Similarity	31.5%	Pred. No. 2e-17;	91	reads 11; Cans 7.

[illegible]

```
QY      60 ISFSLAVIDPAVHL_1SWAGAYIWLANKIATSPTKEDGLOAK--PISFWASFAIOF 115  
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
| : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
```

Db 71 1A-TLLIYR-ALHIGKIIGSYL-CWLMWKTATASYORLDISSKQVTAFOGGLQF 128
Qy 116 UNVXILLYGTALSTFVLPTQALSWVGVSVLLAMIGTFRCNVCMALAGHLFQGLPEYXG 175
Db 129 LNPBAMMGAGVGSFSLAGDYGISGIVSLVNLVPLAGWMLIGGFTRIS-FLQSR 187
Qy 176 RQ--LNIVLALLVYC 189
Db 188 RAMELFNIMIGILTMAC 204

RESULT 6
US-09-328-352-7122
; Sequence 7122, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7122
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7122

Query Match 19.2%; Score 191.5; DB 4; Length 238;
Best Local Similarity 27.1%; Pred. No. 3,18-13;
Matches 52; Conservative 46; Mismatches 79; Indels 15; Gaps 8;
Qy 6 LSAFWTYTLITAMTGPNNIILSSATSHGPROSTRVLAKMSLGLIYMLICAGISPSLA 65
Db 41 LTAFSLSFAIVASIRPGPTNFIILSSHYKISKTLPVILGSCIGALLVIV-VGIGLGT 99
Qy 66 VID-PAVHLLSWGAAYIVWLAKIAT-SP--TKEDGLQAKPISFWASPALQPVNKKII 121
Db 100 ILAVPVIGKIMASGLIMLTVALKLYNNVNVISLEKNEQYPPIGFAAFIMQAINPKTW 159
Qy 122 LYGVATLSTFVLPTQALSWVGVSVLLAMIGTFGNVC---WALAGHLFORLFRQYGRQL 178
Db 160 MMAFAVISVYT---KQODILVNVSIISCIFLILAFCLVYMALVGNLSTRLSK-PKH 215
Qy 179 NI---VIALILV 187
Db 216 NIEKIMAILL 227

RESULT 7
US-09-328-352-7159
; Sequence 7159, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7159
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7159

Query Match 17.9%; Score 178; DB 4; Length 241;
Best Local Similarity 24.6%; Pred. No. 1e-11;
Matches 56; Conservative 48; Mismatches 78; Indels 46; Gaps 10;
Qy 1 VPTLLSAF-----WTYTL-----ITAMTGPNNIILSSATSHGPROSTRVL 43

Db 17 MNPYILHFIPIRMTYTWALILPYLIATLTLLTPGLDTLLIRATLEGSKAPQAA 76
Qy 44 AGMSLGLI--VMLLCAGISFSLAVIDPAVHLLSWGAAYIVWLAKIATSPK----- 96
Db 77 LGISLGIAMGIYVAC-GIG-ALLMASDLAFNLIKMGAAIYMLGIMLKRPQSOLADI 134
Qy 97 EDGLQAKPIS---FMASFLAQFVNKKIILYGVATLSTFVLPTQALSWVGVSVLLAMIG 153
Db 135 QDNHNSRNSISENMFIFGFGNNILNPKVGIYISFLPQFIIPAQSAVYTWGIVHIVIG 194
Qy 154 TFGNVCWA-----LAGHLFORLFRQYGRQL---VIALILVY 188
Db 195 ----VLMSSULLIAMQPLSRYLKQPKVKYMDRITGSIIFLPAKLKAF 238

RESULT 8
US-09-252-991A-32455
; Sequence 32455, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32455
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32455

Query Match 17.4%; Score 173; DB 4; Length 271;
Best Local Similarity 27.9%; Pred. No. 4.3e-11;
Matches 56; Conservative 36; Mismatches 79; Indels 30; Gaps 7;
Qy 5 LLSAFWYTLITAMTGPNNIILSSATSHGPROSTRVLAKMSLGLIYMLICAGISFSL 64
Db 73 LLLAMGAFSLSLISSEPNVLTIVASGANGHFRKTLPEYGVATLGF-VLLAVGTFVR 131
Qy 65 AV-IDPAVHLLSWGAAYIVWLAKIATSPK---EDGLQAKPISFMASFLAQFVNKK 119
Db 132 AIEAYPRFPDYLGAGAAFIAGVRIATADPRLALENGVP---GFGGVLLQWLNPK 187
Qy 120 IILYGVATLSTFVLPTQALSWVGVSVLLAMIGTFGNVC---WALAGHLF----- 167
Db 188 AMIACASGVALFASFSFTHA-----PLLVFAIYLVVCYLSLAAMALLGDRVALLDS 239
Qy 168 QRLFRQYGRQNTIVLALLVY 188
Db 240 PRVRLPNRPMGTLVITAGY 260

RESULT 9
US-09-328-352-8249
; Sequence 8249, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 8249
; LENGTH: 235

```

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18746
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18746

Query Match          15.9%; Score 158.5; DB 4; Length 206;
Best Local Similarity 27.0%; Pred. No. 1,2e-09;
Matches      51; Conservative 35; Mismatches 98; Indels 5; Gaps 4;

QY      6 LSAFTYTLITAMTPGPNNTIALSSATSHGFROSTFVLGMSLGF-LIYMLCAGISFSL 64
        | : : : | | | | | : : : | : : : | : : : | : : : | : : : |
DB      13 LVPFLFPAVVASITPGPTNIIIVLINSQRGLAAMPFIVLGACAVALIILILGIGL-EL 71
        | : : : | | | | | : : : | : : : | : : : | : : : | : : : |

QY      65 AVIDPAAHLLSWAGAAIYIWLAWKIATSPTEKDGQ-AKPISFMASFALQFNVKILY 123
        | : : : | | | | | : : : | : : : | : : : | : : : | : : : |
DB      72 LRRHLLQOGLAMLVGGWLSYLAWSLFRSAGGIDGAPPRLDVLGALLQVNPXAMM 131
        | : : : | | | | | : : : | : : : | : : : | : : : | : : : |

QY      124 GVTALSTFVLPTOTALSWWVGVSILLAMIGTFGNVCMALAGHLFQRLFRQYG--ROLNIV 181
        | : : | | : : : | : : : | : : : | : : : | : : : | : : : |
DB      132 ALAALALPAGBAGAGAGRIGLLALFLFVVSPLCLASMLLGVGSANLLRPSLMKRNQG 191
        | : : | | : : : | : : : | : : : | : : : | : : : | : : : |

QY      182 LALLLVYCA 190
        | : : : |
DB      192 MALLLLASA 200
        | : : : |

RESULT 12
US-09-328-352-6025
; Sequence 6025, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6025
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6025

Query Match          15.6%; Score 155.5; DB 4; Length 214;
Best Local Similarity 22.5%; Pred. No. 2.7e-09;
Matches      45; Conservative 49; Mismatches 89; Indels 17; Gaps 5;

QY      5 LLSAFTYTLITAMTPGPNNTIALSSATSHGFROSTFVLGMSLGLIYMLCAGISFSL 64
        | : : : | | | | | : : : | : : : | : : : | : : : | : : : |
DB      16 MFASMFLEFSFAMSTIPGPVNTYILSTSLNGLRSLPYISGATIGFTLLIFMAFGLOS 75
        | : : : | | | | | : : : | : : : | : : : | : : : | : : : |

QY      65 AVIDPAAHLLSWAGAAIYIWLAWKIATSPTEKDGLOAKPI-----SFMASFALQFNV 118
        | : : : | | | | | : : : | : : : | : : : | : : : | : : : |
DB      76 LTQFPVLVLIKILAVCGTLFCYIGIKIILS-AAIISISSVEVEQMIIINFPDGFLLQWLN 134
        | : : : | | | | | : : : | : : : | : : : | : : : | : : : |

QY      119 KIIL--YGVTAISTFVLPTOTALSWWVGVSILLAMIGTFGNVCMALAGHLFQRLFRQYG 175
        | : : : | | | | | : : : | : : : | : : : | : : : | : : : |
DB      135 KAMLCVSGITMFTSIENQSLPIFIIFYFTCYACL-----FFWGLCGDKFSVVLNQN 189
        | : : : | | | | | : : : | : : : | : : : | : : : | : : : |

QY      176 R--QANIYVALILVCAVRI 193
        | : : : |
DB      190 RUKYFIILMGAFILISALRI 209
        | : : : |

```

```

: Sequence 27175, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: PRIORITY FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 27175
: LENGTH: 237
: TYPE: PR1
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-27175

```

Query Match	15.6%;	Score 155.5;	DB 4;	Length 237;
Best Local Similarity	30.4%;	Pred. No. 3.2e-09;		
Matches 51; Conservative	33;	Mismatches 63;	Indels 21;	Gaps 9;

Oy 3 PTTLSAFTYTL---ITAMTGGNNIIIALSSATSHGFRQSTRVLAGMSIGFLTYMLCA 58
 - - - | : ||| ::||::| : : :
Db 23 PTLGITDWTYYLGVFVLLPGNSLFVALTSRQRGVAITYRACGVFELGDAAVMILSA 82

Oy 59 -G-SFSLAVIDPRAVHLISWAGAAIVYVW-----AMKATSPKEDGLQAKPIS---- 106
 Db 83 LGVA-SLKAEPMLFIGLKYLGAAVLYFYGVGMKRGAMKRLRN- EATAAQAEEKVDHOP 140

Oy 107 PMSFALGVNVKRIILYGTATLSTFLVLP--QTQALSMTWVGVSILLAMI 152
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 141 FRKKLLLSISNPKAILFFETISFFIOFVDGVAYPGGISPLV-LAVIILELV 187

RESULT 14

```

US-09-328-352-6808
; Sequence 6808, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Brelton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAYMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6808
; LENGTH: 210
; TYPE: PR1
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6808

```

Query Match	15.5%	Score 154;	DB 4;	Length 210;
-------------	-------	------------	-------	-------------

Best Local Similarity: 23.6%; Pred. NO. 3.9e-09;
Matches 48; Conservative 53; Mismatches 78; Indels 24; Gaps 7;

QY 6LSAFWYTLITAMPGPNNIIALLSSATSISGFROGTRVLAGMSLG-FLIYMLLCAGISFSL 64

```
QY      65 AVIDPAAVHLLSNMGAAYIVLWAMK--IATSPK-----EGLQAKPISFASFALQ 114
```

QY 115 FVAVKILILXGVTALSTFVLPTQTALSMWVGVSVLAMICTGFANCMWA-----LAG 164

QY 165 HLFORLFRQYGRQLNIVALLLV 187
: | : | : | : ||| : || :

Db 187 NNFSLNKVSG-CIFLLAVLLI 208

RESULT 15
US-09-252-991A-20747

```

Sequence 20747, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
Applicant: Marc J. Rubenfield et al.
Title of Invention: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
Title of Invention: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
File Reference: 107196.136
Current Application Number: US/09/252,991A
Current Filing Date: 1999-02-18
Prior Application Number: US 60/074,788
Prior Filing Date: 1998-02-18
Prior Application Number: US 60/094,190
Prior Filing Date: 1998-07-27
Number of Seq ID NOS: 33142
Seq ID NO 20747
Length: 211
Type: PRT
Organism: Pseudomonas aeruginosa
IS-09-252-991A-20747

```

Query Match	15.4%	Score	153.5	DB	4	Length	211
Best Local Similarity	28.4%	Pred.	No. 4.5e-09				
Matches	50	Conservative	28	Mismatches	75	Indels	23
						Gaps	7

QY 18 VTTPGPNIIILASSATSIGFROSTRVLAGMSGFELIWLWLCAGISFSIAVIIDPAHVHLSM 77
:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 21 LAFPPNNLLSNNAARHGFAATSLAGGRLLAPAGMIATAA--GLPLV---LHTSAW 73

QY -----AGAAIYIVWLA---KXIATSPPTKBDLQAKRISFMAFSALQFV-----NVKILTY 123

78 |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 123

D6 LFLAIKVLGAAYLLWLAIVQLMRTDAQDLVNESPARP-SLMTSLGRQSFLLVAAGNPKAILTI 132

74 LFLAIKVLGAAYLLWLAIVQLMRTDAQDLVNESPARP-SLMTSLGRQSFLLVAAGNPKAILTI 132

QY 124 VTALSTFEVLPTQTALSMWVGSVLLMIGFEGNVCALAG-HLPSRLFRQYGRQL 178
Db 123 FTAFLEQFVDDPGPLGKQPAQAGAAFLLEMLATIALYSYAGHLHGLRIAGQRARRL 188

Search completed: September 10, 2005, 01:46:43
Job time : 43 secs

This Page Blank (uspto)

Db 121 ILVGTALSTFVLPGQALSWVGVSVLTMIGTFGVNCALAGHLFORLFRQGRQNT 180
QY 181 VLALLVYCAVRIFY 195
Db 181 VLALLVYCAVRIFY 195

RESULT 2

US-10-282-122A-60157
; Sequence 60157, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60157
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae

US-10-282-122A-60157
Query Match 16.2%; Score 161.5; DB 15; Length 206;
Best Local Similarity 27.0%; Pred. No. 5e-08;
Matches 53; Conservative 41; Mismatches 81; Indels 21; Gaps 9;
QY 8 AFMTYTLTAMTPGPNNILASATSHGFROSTVLGMSGLFLVLMCA---GISFSL 64
Db 8 AYLLTSLISLSPSGAINTMTTISINHGVRGAASISAGLQGLVHIVLVGVGLGTLFSR 67
QY 65 AVIDPAVHLISWGAAYIWMIA---WKIATSPTEKED--GIQAKPISFWASFAIQFVNVK 119
Db 68 SVL---AFEVLMKGAAYILWLGIOQRPAAIDAINTLAKQOTRGKLFQRAVFVNLNPK 124
QY 120 IILVGTALSTFVLPGQALSW-VGVSV---LLAMIG--TFGN--VCNALAGHLFOR 169
Db 125 SIVFLALFPGFILPHQPGVWQVYLVLGVTTIVVDIIVMIGVATLAQRISATIKSPKQKKA 184

QY 170 LFRQGRQNTIVLAL 185
Db 185 LNRVFGSLFMVLGALL 200

RESULT 3

US-10-282-122A-67883
; Sequence 67883, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67883
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Pseudomonas putida

US-10-282-122A-67883
Query Match 15.9%; Score 158.5; DB 15; Length 211;
Best Local Similarity 26.3%; Pred. No. 1e-07;
Matches 52; Conservative 41; Mismatches 82; Indels 23; Gaps 9;
QY 9 FMTYTLTAMTPGPNNILASATSHGFROSTVLGMSGLFLI-VMLCAGISFSLAVI 67
Db 10 FFAACWVSLISLSPGAIASWSSGLQYGFWRGYNNAIGLQGLIMQAIITIAAGVAVLAA- 68
QY 68 DPAVHLISWGAAYIWMIA---IATSPTEKEDIQ--AKTISF-WASFAIQFVNVK 120
Db 69 SATAFQVTKMFQGVYLVLYLVKQWRALPMDMSDESGVRPIGKPLSVFRGFLVINSNPKA 128
QY 121 IILVGTALSTFVLPGQALSWVGVSV-----LLAMIGTFGVNCALAGHLFORLFR-- 172
Db 129 LVFMALVLPQFLPHNPAHLRQYVAITVTVVDLVLMAGITG-----LASHVL-RHLKTP 182
QY 173 QYGRQNTIVALLLVYCA 190


```

Sequence 10918, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHITUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ. ID NOS: 15109
SEQ ID NO 10918
LENGTH: 217
TYPE: PRT
ORGANISM: Streptomyces avermectilis
US-10-156-761-10918

Query Match          14.3%; Score 142; DB 14; Length 217;
Best Local Similarity 28.7%; Pred. No. 5e-06;
Matches 52; Conservative 26; Mismatches 71; Indels 32; Gaps 8;

QY 3 PTLTSAFWYTTTLTAMTPGPNNTLALSSATSHGPROSTRVLAGMSGFLVLMTCAGISF 62
DB 8 PTLTAGL---VLTVLTPGPNLSLVLSVARKKGVRAGYTAAGVCGDVTVMTLSAAGVA 63

QY 63 SLAVIDPAVAHLLSMWAGAYIVWL-----ANKI-ATSEPKEDGLQA--KPIS----FW 108
DB 64 SLQGANALLFGYIKYKAGAGVLTWLARGMLAAEMEMRTTRDRDAAAPVAPVAGEPFR 123

QY 109 ASFALQFVNKILLYGVTLSTFVL-P--QTQALSWVYGVSVLAMIGTPGNVCMLAAGHL 166
DB 124 RAEVVSLEFNKALTLFVAFVQVDPGAYPALSPFV-----LGAFAQ-----LASPL 171

QY 167 F 167
DB 172 Y 172

RESULT 10
US-10-282-122A-76964
Sequence 76964, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EUTIRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26

```

```

; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 76964
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-10-282-122A-76964
```

```

Query Match      13.8%; Score 137.5; DB 15; Length 205;
Best Local Similarity 28.0%; Pred. No. 1.3e-05;
Matches 53; Conservative 30; Mismatches 77; Indels 29; Gaps 11;
```

```

QY      10 WTYTLITLH---MTPGNNILASSATSHGFROSTRVLAKMSLGLFLVMLLCAGISFSLA 65
           |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      6 WLAVLTAIVVFSIAPGGVTSISNGSYGRHSLGAIIGIQIG-LACHIVLVGIGIGAL 64
QY      66 VIDPA-VVHLISWAGAIVWLAA---WK---IATSPTEKEDGQAKSISFWASFLQFVN 117
           |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      65 VAGALAAFLILKMGAVLWLGIGQKRDAPLTATTSHLSQALLR--KAVLIVLTVN 122
QY      118 VKIILYGVTAISTFVLDPQALSN---VVGV-SVLLAMIGTFGNVCMALAGHLFORLFR 172
           |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      123 PKSIIVFALFPQGIIDPTRD--HWPQFLVIGITVTITDAIVMFGYT--ALAAQL-----G 173
QY      173 QYGRQLNIV 181
           |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      174 RYIRSPNIM 182
```

```

RESULT 11
US-09-927-395-2
```

```

; Sequence 2, Application US/09927395
; Patent No. US20020058314A1
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA
; APPLICANT: ALCOSHIN, VLADIMIR VERNYAMOVICH
; APPLICANT: BELAREOVA, ALL VALENTINOVNA
; APPLICANT: TOKHMAKOVA, IRINA LVOVNA
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L0-HOMOSERINE AND METHOD
; TITLE OF INVENTION: FOR PRODUCING L-AMINO ACIDS
; FILE REFERENCE: 0010-1039-0
; CURRENT APPLICATION NUMBER: US/09/927,395
; CURRENT FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 09/396,357
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: RU98118425
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-927-395-2
```

```

Query Match      12.7%; Score 126; DB 9; Length 205;
```

```

Best Local Similarity 24.5%; Pred. No. 0.00019;
Matches 50; Conservative 36; Mismatches 80; Indels 38; Gaps 10;

QY      8 AFWTYTLITMTGPNNILASSATSHGF-----RQSTRVLAKMSLGLFLVMLL 56
           |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      8 AYLTLSIILTLSPGSAINTMTTSLNHGYPAGGVTCMASDRIGDSYCAQNR----- 58
QY      57 CAGISFSLAVIDPAVHLISWAGAIVWLAA---WKIATSPTEKED--GLQAKSISFWASF 111
           |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      59 GVGTLFSRSVI---AFEVLMKAGAVLWLGIGQWRRAAGAILDKSLASTQSRRLHFORAV 115
QY      112 ALQFVNVKIILYGVTAISTFVLDP-QTQALSMTV-GYSV---LLAMIG--TFGN--VCMA 161
           |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      116 FVNLTPKSIIVFALALFPQFIMPQPOLMQYIVLGVTTIVVDIIVMIGVATLAORIALMT 175
QY      162 LAGHLFORLFPQYGRQLNIVLALL 185
           |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      176 KGPQKMKALNKIFGSLFPLVGAALL 199
```

```

RESULT 12
US-09-847-392-2
```

```

; Sequence 2, Application US/09847392
; Patent No. US20020102670A1
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA
; APPLICANT: ALCOSHIN, VLADIMIR VERNYAMOVICH
; APPLICANT: BELAREOVA, ALL VALENTINOVNA
; APPLICANT: TOKHMAKOVA, IRINA LVOVNA
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L0-HOMOSERINE AND METHOD
; TITLE OF INVENTION: FOR PRODUCING L-AMINO ACIDS
; FILE REFERENCE: 0010-1039-0
; CURRENT APPLICATION NUMBER: US/09/847,392
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 09/396,357
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: RU98118425
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-847-392-2
```

```

Query Match      12.7%; Score 126; DB 9; Length 205;
Best Local Similarity 24.5%; Pred. No. 0.00019;
Matches 50; Conservative 36; Mismatches 80; Indels 38; Gaps 10;
```

```

QY      8 AFWTYTLITMTGPNNILASSATSHGF-----RQSTRVLAKMSLGLFLVMLL 56
           |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      8 AYLTLSIILTLSPGSAINTMTTSLNHGYPAGGVTCMASDRIGDSYCAQNR----- 58
QY      57 CAGISFSLAVIDPAVHLISWAGAIVWLAA---WKIATSPTEKED--GLQAKSISFWASF 111
           |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      59 GVGTLFSRSVI---AFEVLMKAGAVLWLGIGQWRRAAGAILDKSLASTQSRRLHFORAV 115
QY      112 ALQFVNVKIILYGVTAISTFVLDP-QTQALSMTV-GYSV---LLAMIG--TFGN--VCMA 161
           |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      116 FVNLTPKSIIVFALALFPQFIMPQPOLMQYIVLGVTTIVVDIIVMIGVATLAORIALMT 175
QY      162 LAGHLFORLFPQYGRQLNIVLALL 185
           |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      176 KGPQKMKALNKIFGSLFPLVGAALL 199
```

```

RESULT 13
US-10-282-122A-66814
```

```

; Sequence 66814, Application US/10282122A
; Publication No. US20040029129A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 76614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66814
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66814

Query Match      12.4%; Score 123.5; DB 15; Length 209;
Best Local Similarity 24.1%; Pred. No. 0.00035;
Matches 47; Conservative 34; Mismatches 97; Indels 17; Gaps 6;

QY      8 AFWTYTLTAMTPPPNNILSSATSHGFROSTRVLAGMSLGF-LIYMLCAGISFSLAV 66
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      8 AFPLACVAISLSPAGAIASMSGLOYGFAFGYNNALGLQIGLALQIAIVAAVGALAT 67
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      67 IDPAAVHLLSWAGAAYIVWLA---WKTATSPTEKEDGLQ--AKPISF-WASPALQFVWVKI 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      68 -SALAFSLIKMFGAIVLVLAVRQWQAPQALSTGDEBPLGRPLTLVRGFLVNASPKA 126
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      121 ILVGVTALESTFVLDPOTQALSW-----VVGVSVLLAMITGFGNVCWALAGHLFQRLF 171
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      127 VIFWLAIVLPQFIDHPHPLAQYLLMGGMIVVDLIWAGYVGLAARVLRVLRSPRQKLV 186
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      172 RQYGRQINIVALLL 186
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      187 NRTFASLFFVGAAGL 201
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-10-282-122A-44873
; Sequence 44873, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
```

```
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 76614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44873
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-10-282-122A-44873

Query Match      12.0%; Score 120; DB 15; Length 181;
Best Local Similarity 26.8%; Pred. No. 0.00067;
Matches 48; Conservative 35; Mismatches 74; Indels 22; Gaps 9;

QY      28 LSSATSHGFROSTRVLAGMSLGFLL-VMLCAGISFSLAVIDPAAVHLLSWAGAAYIVWL 86
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1 MSSGLNTGFHGHYNNALIGLQIALIQLMTIVAAVGVLPAF-TPLAFQAVKMFVAVLYL 59
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      87 AMKIATSPTEKEDGLQ---AKPIS--FWASPALQFVWVKIILVGVTALESTFVL-LPOTQAL 139
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      60 AYLQWTPAPVDIRIQRHKKKXSVALINGFVNINSPKAIIVFLVATLPQFLDSKQWMI 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      140 SWVGVSV---LMAITGFGNVCWALAGHLFQRLFRQYGRQ--INTVALLIVCAV 191
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      120 QYLIMATWVTIDLIWAGYTG-----LASKVL-RLRSPKQKYLNRGFAVWFSNAL 172
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
US-10-282-122A-69338
; Sequence 69338, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyekind, Judith
; APPLICANT: Wall, Daniel
```

```
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See file wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 69338
/ LENGTH: 210
/ TYPE: PRT
/ ORGANISM: Pseudomonas syringae
US-10-282-122A-69338
```

```
Query Match 11.8% Score 117.5; DB 15; Length 210;
Best Local Similarity 22.6%; Pred. No. 0.0014;
Matches 45; Conservative 40; Mismatches 89; Indels 25; Gaps 9;
```

```
QY 9 FWTYTLITAMTPGPNNTILALSATSHTSGFROSTRVLAGMSIGFLI-VMLCAGISPSLAVI 67
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 9 FFAACWLIISLSPGAGALASNSCGLYGFLRGYNALGIQIALVAQIAIIVAAGLAVLAAS 68
QY 68 DPAAVHLISWAGAIYIWLAA--WKIATSPTK-EDGIQAKPIS-----FWNSPILQFVNV 118
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 69 E-MAFTLIKWFGVAVLYLVIGIKQWR--ASPTDLADESAVRPVGKPMTLVFRGFLVNI 125
QY 119 KIILYGVTAALSTFVLPGQALSWVVGSVLLAMIGTGNVCMALAGH--LFGRLFR---- 172
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 126 KALFIFLAILPQFLPTAPLPMQYVITIAATWVVDLI-----VWAGITGLASKVLRALK 180
QY 173 -QYGRQLNIVLALLLVYCA 190
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 181 PROQRRLNRTFASLFGVGA 199
```

Search completed: September 10, 2005, 01:59:10
Job time : 170 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2005, 01:35:41 ; Search time 40 Seconds

(without alignments)
469.057 Million cell updates/sec

Title: US-10-620-487-2

Perfect score: 996

Sequence: 1 VPTLLSAFWTYTLITAMTP.....RLNIVLALLVYCAVRIFY 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	993	99.7	195	2	A65036	hypothetical 21.2
2	979	98.3	195	2	D91059	hypothetical prote
3	979	98.3	195	2	B85904	hypothetical prote
4	884	88.8	195	2	AF0830	probable membrane
5	272.5	27.4	205	2	B83356	conserved hypotet
6	235.5	23.6	200	2	A10022	probable ABC-trans
7	172	17.3	204	2	B82166	conserved hypotet
8	169	17.0	222	2	B87264	efflux protein, Ly
9	165.5	16.6	206	2	B91223	RhtB family transp
10	164	16.5	205	2	A12614	dihydroadipicinat
11	164	16.5	224	2	H97396	efflux protein, Ly
12	163	16.4	210	2	B87252	homoserine/homoser
13	162.5	16.3	249	2	AG3485	conserved hypotet
14	160.5	16.1	222	2	B82353	hypothetical prote
15	158.5	15.9	197	2	B83280	hypothetical prote
16	158.5	15.9	216	2	C55580	hypothetical prote
17	158	15.9	210	2	D84016	hypothetical prote
18	156.5	15.7	206	2	AF0917	homoserine/homoser
19	156.5	15.7	206	2	AG0466	probable homoserin
20	156.5	15.7	235	2	AD2904	RhtB family transp
21	156.5	15.7	235	2	F97679	hypothetical prote
22	155.5	15.6	209	2	B82388	conserved hypotet
23	155.5	15.6	216	2	F83051	conserved hypotet
24	154.5	15.5	207	2	D83187	hypothetical prote
25	153	15.4	212	2	H87498	LySE family transp
26	143.5	14.4	212	2	F64940	hypothetical prote
27	143	14.4	273	2	AF2805	homoserine/homoser
28	143	14.4	273	2	F97584	hypothetical prote
29	142.5	14.3	212	2	G85790	hypothetical prote

30	140.5	14.1	212	2	C90942	hypothetical prote
31	139	14.0	212	2	AD0714	probable membrane
32	137.5	13.8	205	2	G82358	conserved hypotet
33	135.5	13.6	278	2	B98301	hypothetical prote
34	134.5	13.5	210	2	AC2982	RhtB family transp
35	130	13.1	207	2	B83703	hypothetical prote
36	129	13.0	208	2	C82471	conserved hypotet
37	123.5	12.4	206	2	C82979	hypothetical prote
38	123.5	12.4	209	2	C82990	hypothetical prote
39	123	12.3	213	2	C82523	amino acid transpo
40	122.5	12.3	205	2	AH3641	transporter, lySE
41	122	12.2	206	2	C82139	conserved hypotet
42	122	12.2	216	2	A97675	amino acid efflux-
43	120	12.0	208	2	G87305	efflux protein, Ly
44	119.5	12.0	204	2	B82410	conserved hypotet
45	119.5	12.0	208	2	G84086	dihydrodipicolinat

ALIGNMENTS

```
RESULT 1
A65036
hypothetical 21.2 kD protein in srmb-ung intergenic region - Escherichia coli (strain K
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: A65036
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: A65036
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-195 <BLAT>
A:Cross-references: UNIPROT:P38101; GB:AB000344; GB:U00096; NID:G1788927; PIDN:AACT5631
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: yfjK

Query Match
Best local similarity 99.7%; Score 993; DB 2; Length 195;
Matches 194; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPTLLSAFWTYTLITAMTPGPNIIILSSATSHGFRQSTRVLAAGMSLGFLLYMLUCAGI 60
:|||||
Db 1 MPTLLSAFWTYTLITAMTPGPNIIILSSATSHGFRQSTRVLAAGMSLGFLLYMLUCAGI 60

QY 61 SFLAVIDPAVAHLLSMAGAAITVLAWKIATSPTEKEDGIAKPISFMSFALQFNVVKI 120
:|||||
Db 61 SFLAVIDPAVAHLLSMAGAAITVLAWKIATSPTEKEDGIAKPISFMSFALQFNVVKI 120

QY 121 ILVGVTAISFVLPOTQALSMVGVSVLLAMIGTFGNVCALAGHLFORLFRQYGRQNT 180
:|||||
Db 121 ILVGVTAISFVLPOTQALSMVGVSVLLAMIGTFGNVCALAGHLFORLFRQYGRQNT 180

QY 121 ILVGVTAISFVLPOTQALSMVGVSVLLAMIGTFGNVCALAGHLFORLFRQYGRQNT 180
:|||||
Db 121 ILVGVTAISFVLPOTQALSMVGVSVLLAMIGTFGNVCALAGHLFORLFRQYGRQNT 180

QY 181 VVALLLVYCAVRIFY 195
:|||||
Db 181 VVALLLVYCAVRIFY 195

RESULT 2
D91059
hypothetical protein ECs3444 [imported] - Escherichia coli (strain O157:H7, substrain R
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: D91059
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A96629; MUID:21156231; PMID:11258796
```

A:Accession: D91059
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-195 <HAY>
A:Cross-references: UNIPROT:O8XA19; GB:BA000007; PIDD:G13362915; GSPDB:G
A:Experimental source: strain O157:H7, substrain R1MD 050952
C:Genetics:
A:Gene: EC63444

Query Match	98.3%	Score 979	DB 2	Length 195
Best Local Similarity	96.9%	Pred. No. 3.3e-80		
Matches 189	Conservative 4	Mismatches 2	Indels 0	Gaps 0
QY	1	VTPTLLSAFWYTYTTITAMTPGPNNILALSSATSHGFROSTVTLAGMSLGIPLVLLCAGI	60	
	:	: : : : : : : : : :		
Db	1	MTPTLLSAFWYTYTTITAMTPGPNNILALSSATSHGFROSTVTLAGMSLGIPLVLLCAGI	60	
QY	61	SFSLAVIDPRAVHLHLSWAGAAIYIWLAKKITSTPKEDGLOAKISTWASALQFVAVKI	120	
	:	: : : : : : : : :		
Db	61	SFSLAVIDPRAVHLHLSWAGAAIYIWLAKKITSTPKEDGLOAKISTWASALQFVAVKI	120	
QY	121	ILVGTALSTFVLPTQALSWWGVSVLLAMIGTFGNVCMAHLFQRLFRQYGRQNTI	180	
	:	: : : : : : : : :		
Db	121	ILVGTALSTFVLPTQALSWWGVSVLLAMIGTFGNVCMAHLFQRLFRQYGRQNTI	180	
QY	181	VLAALLVTCAVRIIFY 195		
	:	: : : : : : : :		
Db	181	VLAALLVTCAVRIIFY 195		

RESULT 3
 B85904
 Hypothetical protein yf1k [imported] - *Escherichia coli* (strain O157:H7, substrain EDL933)
 C.Species: *Escherichia coli*
 C.Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
 C.Accession: B85904
 R.Perna, N.T., Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apodaca
 Nature 409, 523-533, 2001
 A.Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A.Reference number: A85480; MUID:21074935; PMID:1120551
 A.Accession: B85904
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-195 <STO>
 A.Cross-references: UNIPROT:O8XA19; GB:AEO05174; NID:g12516999; PIDN:AA657694.1; GSPDB:G
 A.Experimental source: strain O157:H7, substrain EDL933
 C.Genetics:
 A.Gene: yf1k

[illegible]

RESULT 4
AF0830

probable membrane protein SRY2838 [imported] - *Salmonella enterica* subsp. *enterica* sero
C|Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
A|Note: This species has also been called *Salmonella typhi*
C|Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C|Accession: AF0830
P|Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar
S.; Moule, S.; O'Gaora, P.
Nature 413, 846-852, 2001.
A|Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A|Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* sero
A|Reference number: AB0502; MUID:21534947; PMID:11677608
A|Accession: AF0830
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-195 <PAR>
A|Cross-references: GB:AL513382; PIDN:CAD02794.1; PID:gl16503803; GSPDB:GN00176
C|Genetics:
A|Gene: SRY2838

[illegible]

RESULT 5
B83356
conserved hypothetical protein PA2306 [imported] - *Pseudomonas aeruginosa* (strain PA01)
C/Species: *Pseudomonas aeruginosa*
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C/Accession: B83356
R/Stover, C.K.; Plam, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adaman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim
: Lorry, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic path
A/Reference number: AB2950; MUID:20437337; PMID:10984043
A/Accession: B83356
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-205 <STO>
A/Cross-references: UNIPROT:Q91IG9; GB:AE004657; GB:AE004091; NID:g9948338; PIDN:AAG056
A/Experimental source: strain PA01
C/Genetics:
A/Gene: PA2306

[illegible]

A;Cross-references: UNIPROT:P27847; GB:BA000007; PIDD:BA38177.1; PIDD:913364220; GSPDB:A;Experimental source: Strain O157:H7, substrain R1MD 050952
C;Genetics:
A;Gene: EC94754

Query Match	16.6%;	Score 165.5;	DB 2;	Length 206;
Best Local Similarity	27.0%;	Pred. No. 1.2e-07;		
Matches 53;	Conservative 42;	Mismatches 80;	Indels 21;	Gaps 9;

```

8 AFWTYLLITAMPGPNNILALSSATSHFQROSTRVLGMSLGLIWMICA---GISFSL 64
|: ::||| |::: |||: |||: |||
8 AYLTSIILSLSPGSAINTMTTSLNNGYRGNAASIAGLQTAIHIVLVGVGLTFSR 67

```

```
68 SVI--APEVLKMWAGAYLLIWIGQWRAGAIDLKSLASTQSRRHLFQRVAVFNLTNPK 124
```

120 ILIYGVIALSTFVL-QTQALSMV--GYSV-----LLAMIG--TFGN--VCNALAGHLFOR 169
::|::||::|::|::|::|::|::|:
125 SIYFLAALLPQFIIMPQQQLMGQYLIVLCTTTIVVDIIWMIGATLAQRIALMTIKGRQMKA 184

```

y      170 LFRQYGRQINIVLALL 185
      | : : | : | | |
b      185 LNKIFGSLFMLVGALL 200

```

RESULT 10

```
rhbB family transporter rhbB [imported] - Agrobacterium tumefaciens (strain C58, Dupont),
;Species: Agrobacterium tumefaciens
;Date: 11-Jan-2002 #sequence_rev1sion 11-Jan-2002 #text_change 09-Jul-2004
```

Wood, D.W.; Senubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, I.; Garg, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCellan, Karp, P.; Romero, P.; Zhang, S.

Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.W.
Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

```
;Accession: AI2614
;Status: preliminary
;Molecule type: DNA
```

```

;Cross-references: UNIPROT:0BU12; GS:AE008668; PIDN:AAL41335.1; PID:G17738647; GSPDB:G
;Experimental source: strain C58 (Dupont)
Genetics:

```

```

/Map position: circular chromosome
Query Match      16.5%; Score 164; DB 2; Length 205;

```

Matches	50;	Conservative	33;	Mismatches	66;	Indels	20;	Gaps	4;
1	VPTLLSAFWYTYLLTAMTPGPNII	ALSSATSHGRQSTRVL	AGMSGLFYIMLLCAGI	60					

1 MTTTLLAATATLFTAAADPGMTAIVARALSGRRETFEMGLVLGDMITL---TGV 57
61 SEELAVIDPDA---AYHLLSNAGAAIVLWLAWKIATSTPTKEGLQAK-----PISPMASFA 112

58 ILGLAFVAQTFQEEAFVULKPGAGAAVLLXYIAMKLMTAGLLPQDLKARKKSTSIPIMSFLSGLL 117

118 ITLGNPKTMLEFVVALVPTLIDIRMIGPSSEYATLLALFEVLMVALLPYI 166

RESULT 11

```
hydroxydipicolinate reductase (AA001519) [imported] - Agrobacterium tumefaciens (strain
/Species: Agrobacterium tumefaciens
/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
Accession: H97396
```

R:Goodnar, B., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Quicollo, B., Goldman, A., Liu, P., Wollam, C., Allinger, M., Doughty, D., Scott, C., Lappas, C., Markelz, B., Science 294, 2323-2328, 2001
A:Rille: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens* Reference number: A97359, MUID:21608551, PMID:11743194

```

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-224 <KUR>

```

C:Genetica:
A:Gene: AGR_C_546
A:Map position: circular chromosome

Query Match	16.5%	Score	164	DB 2	Length	224			
Best Local Similarity	29.6%	Pred. No.	1.8e-07						
Matches	50	Conservative	33	Mismatches	66	Indels	20	Gaps	4

```

Oy      1 VTPTLLSAFWYTLITAMTPGGNNILASSATSHGFRGSTRVLAGMSLGFLVMLLCAGI 600
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      20 MTLTLLAYATLFAAALPGSGMTAIVARALSGFRFEFMGGLVLMGIYLV---TGV 760

```

```

QY      61 SFSLSAVIDA---AVHLLSNWAGANIIVMLAMKIATSPTEKEDLQAK----PISFMASFA 112
      || : : : |||||::: || : : : ||||| : : : || : : :
Db      77 ILGLAFVAQTFQGEAFMVLKFKAGAAVLIIYIMKLMTWAGLLPQDLKARKSTSIEMSFLEGGLL 136

```

```

QY      113 LQFVNKILXGVTALSTFV-----LPQIQALSNWGVSVLLAMI 152
      :  |  :  |  :  |  :  |  |  |  |  |  |  |  |  |  |  |
DB      137 ITLGNPKMTLEFYVALPPTLIDIRMGPSSEYATLALTTFVVLNAVLLPYI 185

```

RESULT 12

efflux protein, *lysB* family [imported] - *Caulobacter crescentus*
C.Species: *Caulobacter crescentus*
C.Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004

R.Nieman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; H
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hatt
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.;

A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: E87252

A: Molecule type: DNA-
A: Residues: 1-210 <STO>
A: Cross-references: UNIPROT: Q9AC37, GB: AE005673, NID: g13421119, PIDN: AAK22017.1, GSPDB: C

A:Gene: CC0029
C:Superfamily: hypothetical protein b1798

Best Local Similarity 29.2%; Pred. No. 2.1e-07;
Matches 52; Conservative 29; Mismatches 77; Indels 20; Gaps 7;

Db 9 AFTVLCSMALTPGPNNLYLVSRISICQGRMAGIVSLIGTAAGF-VVYLKCAALGITALLM 67

```
Db      | : | : | : | : | : | : |  
68 AAPAYDLEIFGALYLAWLAWQ-AIRPGASPFQVRELPKDSPAKLITMGFVTMLNDPK 126
```

Db 127 AAMLYISLLPQFIKPEHGVFTQSLA-LGLTQICVSLTVNGVICLAAGTIAGFLAAR 182

RESULT 13

RESULT 13
AG3485
homoserine/homoserine lactone efflux protein [imported] - *Brucella meliensis* (strain 1
C;Species: *Brucella meliensis*

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2005, 01:34:51 ; Search time 173 Seconds

(without alignments)
577.199 Million cell updates/sec

Title: US-10-620-487-2

Perfect score: 996

Sequence: 1 VPTPLLSAFWYTLITAMP.....RLNIVLALLVYCAVRIF 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

UniProt 03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	993	99.7	195	1	YPIK_ECOLI	P38101 escherichia
2	989	99.3	195	2	Q8FPI1	Q8FPI1 escherichia
3	987	99.1	195	2	Q83K22	Q83K22 shigella fl
4	979	98.3	195	2	Q8XA19	Q8XA19 escherichia
5	884	88.8	195	2	Q8Z4U7	Q8Z4U7 salmonella
6	884	88.8	195	2	Q8ZMK5	Q8ZMK5 salmonella
7	367.5	36.9	200	2	Q7P2H4	Q7P2H4 fusobacteri
8	316.5	31.8	194	2	Q63B00	Q63B00 bacillus ce
9	311.5	31.3	194	2	Q81Q08	Q81Q08 bacillus an
10	311.5	31.3	194	2	Q6H1R1	Q6H1R1 bacillus th
11	310.5	31.2	194	2	Q738B5	Q738B5 bacillus ce
12	303.5	30.5	194	2	Q81DI7	Q81DI7 bacillus ce
13	293	29.4	192	2	Q65N28	Q65N28 bacillus li
14	277.5	27.9	198	2	Q829T1	Q829T1 bacillus sp
15	272.5	27.4	205	2	Q911G9	Q911G9 pseudomonas
16	262.5	26.4	209	2	Q88HP4	Q88HP4 pseudomonas
17	258.5	26.0	202	2	Q872H7	Q872H7 pseudomonas
18	256	25.7	202	2	Q7N9S0	Q7N9S0 photobacteri
19	244	24.5	204	2	Q6LKH4	Q6LKH4 photobacteri
20	237.5	23.8	200	2	Q6CZU2	Q6CZU2 erwinia car
21	236	23.7	219	2	Q7W4S0	Q7W4S0 bordetella
22	236	23.7	219	2	Q7WFK7	Q7WFK7 bordetella
23	235.5	23.6	197	2	Q62NS0	Q62NS0 rhizobium m
24	235.5	23.6	200	2	Q664G6	Q664G6 yersinia ps
25	235.5	23.6	200	2	Q8ZJDP	Q8ZJDP yersinia pe
26	234.5	23.5	192	2	Q8CK91	Q8CK91 yersinia pe
27	232.5	23.3	202	2	Q89G49	Q89G49 bradyrhizob
28	232	23.3	219	2	Q7VUS5	Q7VUS5 bordetella
29	231	23.2	196	2	Q7QMP4	Q7QMP4 anopheles g
30	222.5	22.3	197	2	Q8EG90	Q8EG90 shewanella
31	221	22.2	193	2	Q88HC2	Q88HC2 pseudomonas

32	215.5	21.6	199	2	Q8D574	Q8D574 vibrio vuln
33	214.5	21.5	188	2	Q882W8	Q882W8 pseudomonas
34	213.5	21.4	198	2	Q98KX8	Q98KX8 rhizobium 1
35	211.5	21.2	200	2	Q6LYD7	Q6LYD7 photobacter
36	211.5	21.2	206	2	Q7NWS8	Q7NWS8 chromobacte
37	206.5	20.7	208	2	Q7NWP5	Q7NWP5 chromobacte
38	206	20.7	195	2	Q7MG87	Q7MG87 vibrio vuln
39	204.5	20.5	248	2	Q7WDP6	Q7WDP6 bordetella
40	202	20.3	195	2	Q8DAN7	Q8DAN7 vibrio vuln
41	202	20.3	205	2	Q6ZHR2	Q6ZHR2 burkholderi
42	202	20.3	205	2	Q63P95	Q63P95 burkholderi
43	199.5	20.0	208	2	Q7W2P8	Q7W2P8 bordetella
44	185.5	18.6	209	2	Q7V575	Q7V575 bordetella
45	183	18.4	194	2	Q87X08	Q87X08 vibrio para

ALIGNMENTS

RESULT 1

ID	YPIK_ECOLI	STANDARD	PRT	195 AA.
AC	P38101			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	25-JAN-2005 (Rel. 46, Last annotation update)			
DE	Hypothetical protein yfik.			
GN	Name=yfik; Ordered locus Names=b2578;			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12;			
RA	Nashimoto H.;			
RT	"Non-ribosomal proteins affecting the assembly of ribosomes in			
RT	Escherichia coli.";			
RL	(In) Niehaus K.H. (eds.);			
RL	The translational apparatus, pp.185-195, Plenum Press, New York			
RL	(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12;			
RA	Nashimoto H., Saito N.;			
RL	Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.			
RP	[3]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MG1655;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RT	Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12.";			
RL	Science 277:1453-1474(1997).			
RL	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12;			
RC	MDL1NE=97349980; PubMed=9205837;			
RA	Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,			
RA	Itch T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,			
RA	Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,			
RA	Ohnita T., Oyama S., Saito N., Saito N., Saito Y., Sivasubram S.,			
RA	Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,			
RT	Yamagata S., Horiiuchi T.;			
RT	"Construction of a contiguous 874-kb sequence of the Escherichia coli-			
RT	K12 genome corresponding to 50.0-68.8 min on the linkage map and			
RT	analysis of its sequence features.";			
RL	DNA Res. 4:91-113(1997).			
RL	[5]			
RP	IDENTIFICATION.			
RP	Rudd K.E.;			

Unpublished observations (Aug-1994).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: Belongs to the rnt family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: D13169; -; NOT_ANNOTATED_CDS.

DR EMBL: D64044; -; NOT_ANNOTATED_CDS.

DR EMBL: U00096; AAC75631.1; -.

DR EMBL: D90886; BAA16464.1; -.

DR PIR: A65036; A65036.

DR ECHOBASE: EB2339; -.

DR EcoGene: EG12445; yf1K.

DR InterPro: IPR001123; LyseB.

DR Pfam: PF01810; LyseB; 1.

DR Complete proteome: Hypothetical protein; Transmembrane.

FT TRANSMEM 8 28 Potential.

FT TRANSMEM 47 67 Potential.

FT TRANSMEM 71 91 Potential.

FT TRANSMEM 142 162 Potential.

FT TRANSMEM 195 AA; 21248 MW; 5F86B828DDDEC090 CRC64;

SO SEQUENCE

Query Match 99.7%; Score 993; DB 1; Length 195;
Best Local Similarity 99.5%; Pred. No. 5.8e-73;
Matches 194; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VPTPLSAFWYTLITAMTPGPNIIILSSATSHGFRQSTRVLGMSLGFILWLLCAGI 60
DB 1 MPTPLSAFWYTLITAMTPGPNIIILSSATSHGFRQSTRVLGMSLGFILWLLCAGI 60

QY 61 SFSILAVIDPAVHLISWAGAAIYIWLAMKIATSPTEKEDGLQAKPISFWASPALQFVNVI 120
DB 61 SFSILAVIDPAVHLISWAGAAIYIWLAMKIATSPTEKEDGLQAKPISFWASPALQFVNVI 120

QY 121 ILVGVTAISTFVLPTQALSWVGVSVLLAMITFGNVCAALAGHLFQRLFRQYGRQINT 180
DB 121 ILVGVTAISTFVLPTQALSWVGVSVLLAMITFGNVCAALAGHLFQRLFRQYGRQINT 180

QY 121 ILVGVTAISTFVLPTQALSWVGVSVLLAMITFGNVCAALAGHLFQRLFRQYGRQINT 180
DB 121 ILVGVTAISTFVLPTQALSWVGVSVLLAMITFGNVCAALAGHLFQRLFRQYGRQINT 180

QY 181 VLALLVYCAVRIFY 195
DB 181 VLALLVYCAVRIFY 195

QY 181 VLALLVYCAVRIFY 195
DB 181 VLALLVYCAVRIFY 195

RESULT 2

Q8F11 PRELIMINARY; PRT; 195 AA.

AC Q8F11; 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein yf1K.
GN Name=yf1K; OrderedlocusNames=33102;
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN 1 [1]
RP SEQUENCE FROM N.A.
RC STRAIN=06:HI / CPT073 / ATCC 700928 / UPEC;
RX MEDLINE=22386234; PubMed=12471157; DOI=10.1073/pnas.252529799;
RA Weich R.A., Burckles E.L., Plunkett G. III, Redford P., Roesech P.,
RA Raacke D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Domeneberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).

DR EMBL: AE016764; AAN81551.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005293; F:lysine permease activity; IEA.
DR GO: GO:0006865; P:amino acid transport; IEA.
DR InterPro: IPR001123; LyseB.
DR Pfam: PF01810; LyseB; 1.
DR Complete proteome: Hypothetical protein.
SO SEQUENCE 195 AA; 21278 MW; 295DD583ADCA8584 CRC64;

Query Match 99.3%; Score 989; DB 2; Length 195;
Best Local Similarity 99.0%; Pred. No. 1.2e-72;
Matches 193; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VPTPLSAFWYTLITAMTPGPNIIILSSATSHGFRQSTRVLGMSLGFILWLLCAGI 60
DB 1 MPTPLSAFWYTLITAMTPGPNIIILSSATSHGFRQSTRVLGMSLGFILWLLCAGI 60

QY 61 SFSILAVIDPAVHLISWAGAAIYIWLAMKIATSPTEKEDGLQAKPISFWASPALQFVNVI 120
DB 61 SFSILAVIDPAVHLISWAGAAIYIWLAMKIATSPTEKEDGLQAKPISFWASPALQFVNVI 120

QY 121 ILVGVTAISTFVLPTQALSWVGVSVLLAMITFGNVCAALAGHLFQRLFRQYGRQINT 180
DB 121 ILVGVTAISTFVLPTQALSWVGVSVLLAMITFGNVCAALAGHLFQRLFRQYGRQINT 180

QY 121 ILVGVTAISTFVLPTQALSWVGVSVLLAMITFGNVCAALAGHLFQRLFRQYGRQINT 180
DB 121 ILVGVTAISTFVLPTQALSWVGVSVLLAMITFGNVCAALAGHLFQRLFRQYGRQINT 180

QY 181 VLALLVYCAVRIFY 195
DB 181 VLALLVYCAVRIFY 195

QY 181 VLALLVYCAVRIFY 195
DB 181 VLALLVYCAVRIFY 195

RESULT 3

Q83K2 PRELIMINARY; PRT; 195 AA.

AC Q83K2; 07COB0.
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Hypothetical protein yf1K.
GN Name=yf1K; OrderedlocusNames=52813, SF2640;
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN 1 [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590; DOI=10.1093/nar/gkf566;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang Y., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441 (2002).
RN 121
RP SEQUENCE FROM N.A.
RC STRAIN=2457T.
RX MEDLINE=22590274; PubMed=12704152;
RX DOI=10.1128/JAI.71.5.2775-2786.2003;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786 (2003).
DR EMBL: AE015279; AAN44136.1; -.
DR EMBL: AE016987; AAP17960.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005293; F:lysine permease activity; IEA.
DR GO: GO:0006865; P:amino acid transport; IEA.
DR InterPro: IPR001123; LyseB.

DR Pfam; PF01810; LysE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 195 AA; 21260 MM; D77DD7B0A152ED16C CRC64;

Query Match 99.1%; Score 987; DB 2; Length 195;
Best Local Similarity 99.0%; Pred. No. 1.8e-72;
Matches 193; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 VTPTLSAFPWTYTLITMTGPNNTILASSATSHGFQSTRTVLAGMSLGFILVLLCAGI 60
   |||
DB 1 MTPPLLSAFPWTYTLITMTGPNNTILASSATSHGFQSTRTVLAGMSLGFILVLLCAGI 60

QY 61 SFSLSAVIDPAVHLISWAGAAIYWLAMKATSPTEKEDGIQAKPISFWASFALQFVAVKI 120
   |||
DB 61 SFSLSAVIDPAVHLISWAGAAIYWLAMKATSPTEKEDGIQAKPISFWASFALQFVAVKI 120

QY 121 ILVGVTAISTFVLPTQALSWVGVSVLLAMIGTFGNVCMALAGHLFQRLFRQYGRQNTI 180
   |||
DB 121 ILVGVTAISTFVLPTQALSWVGVSVLLAMIGTFGNVCMALAGHLFQRLFRQYGRQNTI 180

QY 181 VLLALLVYCAVRIFY 195
   |||
DB 181 VLLALLVYCAVRIFY 195
```

RESULT 4

08XA19 PRELIMINARY; PRT; 195 AA.
AC 08XA19; 07ABJ3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein yfik (Hypothetical protein EC63444).
GN Name=yfik; Ordered locus names=EC63444, 23861;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NX NCBI_TaxID=83334;

RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RC MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Karpstrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grodeck E.V., Davis N.W., Lim A., Dimalanta B.T., Peticolausis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952 / EHEC;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
DR EMBL; AF005488; AAC57694.1; -;
DR EMBL; AF002562; BAB36867.1; -;
DR PIR; B85904; B85904.
DR PIR; D91059; D91059.
DR CO; GO:0016020; C:membrane; IEA.
DR CO; GO:0005293; F:lysine permease activity; IEA.
DR CO; GO:0006865; P:amino acid transport; IEA.
DR InterPro; IPR001123; LysE.
DR Pfam; PF01810; LysE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 195 AA; 21301 MM; 54DA47EF23488F07 CRC64;

Query Match 98.3%; Score 979; DB 2; Length 195;
Best Local Similarity 96.9%; Pred. No. 7.9e-72;
Matches 189; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 VTPTLSAFPWTYTLITMTGPNNTILASSATSHGFQSTRTVLAGMSLGFILVLLCAGI 60
   |||
DB 1 MTPPLLSAFPWTYTLITMTGPNNTILASSATSHGFQSTRTVLAGMSLGFILVLLCAGI 60

QY 61 SFSLSAVIDPAVHLISWAGAAIYWLAMKATSPTEKEDGIQAKPISFWASFALQFVAVKI 120
   |||
DB 61 SFSLSAVIDPAVHLISWAGAAIYWLAMKATSPTEKEDGIQAKPISFWASFALQFVAVKI 120

QY 121 ILVGVTAISTFVLPTQALSWVGVSVLLAMIGTFGNVCMALAGHLFQRLFRQYGRQNTI 180
   |||
DB 121 ILVGVTAISTFVLPTQALSWVGVSVLLAMIGTFGNVCMALAGHLFQRLFRQYGRQNTI 180

QY 181 VLLALLVYCAVRIFY 195
   |||
DB 181 VLLALLVYCAVRIFY 195
```

RESULT 5

08Z4J7 PRELIMINARY; PRT; 195 AA.
AC 08Z4J7; 07CBN2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Putative membrane protein.
GN Name=yfik; Ordered locus names=STY2838, t0265;
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
NX NCBI_TaxID=601;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jorgels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gara P., Perry K.,
RA Quail M.A., Rutherford K.M., Simmonds M., Skellern J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18";
RL Nature 413:848-852(2001).

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
DOI=10.1128/JB.185.7.2330-2337.2003;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
J and CT18";
RL J. Bacteriol. 185:2330-2337(2003).
DR EMBL; AL627275; CAD02794.1; -;
DR EMBL; AL616835; AAO67991.1; -;
DR CO; GO:0016020; C:membrane; IEA.
DR CO; GO:0005293; F:lysine permease activity; IEA.
DR CO; GO:0006865; P:amino acid transport; IEA.
DR Pfam; PF01810; LysE; 1.
KW Complete proteome.
SQ SEQUENCE 195 AA; 21334 MM; EF30142B85955B85 CRC64;

Query Match 88.8%; Score 884; DB 2; Length 195;
Best Local Similarity 83.6%; Pred. No. 4.1e-64;
Matches 163; Conservative 19; Mismatches 13; Indels 0; Gaps 0;

```
QY 1 VTPTLSAFPWTYTLITMTGPNNTILASSATSHGFQSTRTVLAGMSLGFILVLLCAGI 60
```

```
DB 1 MTMMLSAFTYTLITRLTGPNNIILASATSHGRQSRVLAGMSLGLVWMLCAGI 60
QY 61 SFSIAVIDPAAVHLSWGAAYIYWLAKIATSPTEKEDGLQAKPISFWASFALQFVNVKI 120
DB 61 AFSIAVIDPAIHLISWGAAYIILMLAKIATSPADEKRPKPVGFVWFSFGLQFVNVKI 120
QY 121 ILVGTALSTFVLPQTOALSWVGVSVLTMIGTFGVCWALAGHLFQRLFRQYGRQNT 180
DB 121 ILVGTALSTFVLPQTOALMWIGVSIILALIGFNVCMALAGHLFQRAFRHGRQNT 180
QY 181 VLALLVYCAVRIFY 195
DB 181 ILALLVYCAVRIFY 195

RESULT 6
Q8ZMX5 PRELIMINARY; PRT; 195 AA.
ID Q8ZMX5
AC 08ZMX5;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Putative transport protein.
GN Name=yfjK; OrderedlocusNames=STM2645;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
ON NCBI_TaxID=602;
RX MEDLINE=1534948; PubMed=11677609; DOI=10.1038/35101614;
RA McClelland W., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Portwiliik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan B., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT L12."
RL Nature 413:852-856(2001).
DR EMBL; AEO08820; AAL21539.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005293; F:lysine permease activity; IEA.
DR GO; GO:0006865; P:amino acid transport; IEA.
KW Complete proteome.
SQ SEQUENCE 195 AA; 21292 MW; B63E83C2FB104EB3 CRC64;

Query Match 88.8%; Score 884; DB 2; Length 195;
Best Local Similarity 83.6%; Pred. No. 4.1e-64;
Matches 163; Conservative 18; Mismatches 14; Indels 0; Gaps 0;

QY 1 VTPPLLSAFWYTLITAMTPGPNIIILASATSHGFQSTFVLGMSLGLIWLTCAGI 60
DB 1 MTPPLLSAFWYTLITAMTPGPNIIILASATSHGFQSTFVLGMSLGLIWLTCAGI 60
QY 61 SFSIAVIDPAAVHLSWGAAYIYWLAKIATSPTEKEDGLQAKPISFWASFALQFVNVKI 120
DB 61 AFSIAVIDPAIHLISWGAAYIILMLAKIATSPADEKRPKPVGFVWFSFGLQFVNVKI 120
QY 121 ILVGTALSTFVLPQTOALSWVGVSVLTMIGTFGVCWALAGHLFQRLFRQYGRQNT 180
DB 121 ILVGTALSTFVLPQTOALMWIGVSIILALIGFNVCMALAGHLFQRAFRHGRQNT 180
QY 181 VLALLVYCAVRIFY 195
DB 181 ILALLVYCAVRIFY 195

RESULT 7
Q7P2H4 PRELIMINARY; PRT; 200 AA.
ID Q7P2H4
```

```
AC Q7P2H4;
DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Transporter, lysE family.
GN Name=FN0676;
OS Fusobacterium nucleatum subsp. vincentii ATCC 49256.
OC Bacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
ON NCBI_TaxID=209882;
RX SEQUENCE FROM N.A.
RA Karpalral V., Ivanova N., Anderson I., Reznik G., Bhattacharyya A.,
RA Gardner W.L., Mikhailova N., Larsen N., D'Souza M., Walunas T.,
RA Haselkorn R., Overbeek R., Kyriides N.;
RA Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RL -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABF0100094; EAA23701.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005293; F:lysine permease activity; IEA.
DR GO; GO:0006865; P:amino acid transport; IEA.
DR InterPro; IPR001123; LysE.
DR Pfam; PF01810; LysE; 1.
SQ SEQUENCE 200 AA; 22782 MW; 463825149BE6DA43 CRC64;

Query Match 36.9%; Score 367.5; DB 2; Length 200;
Best Local Similarity 40.8%; Pred. No. 3.6e-22;
Matches 78; Conservative 37; Mismatches 67; Indels 9; Gaps 3;

QY 5 LLSAFWYTLITAMTPGPNIIILASATSHGFQSTFVLGMSLGLIWLTCAGISFSL 64
DB 8 MLESEYPAALITPFGPNIIILASATSHGFQSTFVLGMSLGLIWLTCAGISFSL 67
QY 65 AVIDPAVHLSWGAAYIYWLAKIATSPTEKEDGLQAKPISFWASFALQFVNVKI 123
DB 68 KKVSDTYQNIIMKXIGLYIFWLMHIFKSPSSILENSPREFLVGLGFIQFVNVKMLY 127
QY 124 GVTALSTFVLPQTOA---LSWVGVSVLTMIGTFGVCWALAGHLFQRLFRQYGRQNT 179
DB 128 GWSISFTFIPYQSALFIFLFGFMSIL---GATLAALWALAGSLFQDLRHRIRIFN 183
QY 180 IVALLLVYCA 190
DB 184 TVMGLILVSA 194

RESULT 8
Q63BD0 PRELIMINARY; PRT; 194 AA.
ID Q63BD0
AC Q63BD0;
DT 25-OCT-2004 (T-EMBLrel. 28, Created)
DT 25-OCT-2004 (T-EMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (T-EMBLrel. 28, Last annotation update)
DE Transporter, lysE family.
GN Name=lysE; ORFNames=BTZK2196;
OS Bacillus cereus ZK.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
ON NCBI_TaxID=288681;
RX SEQUENCE FROM N.A.
RA Brettn T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RT "Complete genome sequence of Bacillus cereus ZK."
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CP000001; AAU18061.1; -.
SQ SEQUENCE 194 AA; 21905 MW; 2600626310681709 CRC64;

Query Match 31.8%; Score 316.5; DB 2; Length 194;
```


Db	185	VCSA181	191
RESULT	11		
ID	Q738B5	PRELIMINARY;	PRT; 194 AA.
AC	Q738B5		
DT	05-JUL-2004	(TREMBLrel. 27, Created)	
DT	05-JUL-2004	(TREMBLrel. 27, Last sequence update)	
DT	05-JUL-2004	(TREMBLrel. 27, Last annotation update)	
DE	Transporter, LysE family.		
GN	OrderedLocustNames=BCE2480;		
OS	Bacillus cereus (strain ATCC 10987).		
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
OX	NCBI_TaxId=222523;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	PubMed:14960714; DOI=10.1093/nar/gkh258;		
RA	Raeko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,		
RA	Shores K.A., Fouts D.E., Tourasse N.J., Anguinoi S.V., Kolonay J.F.,		
RA	Neilson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;		
RT	"The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic		
RT	adaptations and a large plasmid related to Bacillus anthracis pXOI.1";		
RL	Nucleic Acids Res. 32:977-988(2004).		
DR	EMBL; AE012772; AAS41397.1; -		
DR	TIGR; BCE2480; -		
DR	GO; GO:0016020; C:membrane; IEA.		
DR	GO; GO:0005293; F:lysine permease activity; IEA.		
DR	GO; GO:0006865; P:amino acid transport; IEA.		
DR	InterPro; IPR001123; LysE.		
DR	Pfam; PF01810; LysE; 1.		
DR	Complete proteome.		
SO	SEQUENCE	194 AA; 21828 MW; DAB92CAE54BE2E26 CRC64;	

RA	Grechkhii Y., Pusch G., Haase Korn R., Fongstein M., Ehrlich S.D.,
RA	"Overbeek R., Kyprides N.C.;
RT	"genome sequence of Bacillus cereus and comparative analysis with
RT	Bacillus anthracis.";
RL	Nature 423:87-91(2003)."
DR	EMBL; AE017005; AAP09341.1; -.
DR	GO:0016020; C:membrane; IEA.
DR	GO:0005293; P:Lysine permease activity; IEA.
DR	GO:0006865; P:amino acid transport; IEA.
DR	InterPro; IPR001123; LysE.
KW	Pfam; PF01810; LysE; 1.
KW	Complete proteome.
SQ	SEQUENCE 194 AA; 21777 MW; 393B66AF87C8C96F CRC64;
Query Match	30.5%; Score 303.5; DB 2; Length 194;
Best Local Similarity	36.2%; Pred. No. 5.5e-17;
Matches 68; Conservative 39; Mismatches 78; Indels 3; Gaps 2;	
Dd	8 AFWTYYLTITAMTPGPNNNIIASATSHGFROSTFVLAMGSLGLIYVWLCCAGISFLSIAVI 67 5 SFLLFVFISSTPPGNPFAMTYANOHQLKRSINFCGVAFGFILTSLSFFENIVLINI 64
Qy	68 DPAAVHLISMGAAYIYVWLAKTIATSP--TKEDGLQAKPISFMASPALOPNVNIIILYGV 125 65 IPILIEFPKLIIGAVMYMLAPKIKITSGGAGAHNKNNKL-FYTGIFLOPIFNPKGILFGL 123
Qy	126 TALSTFVLDPOTQALSWVVGVSVLLAMITGTFGNVCWALAGHLEQRLFRQYGRLNIIVALL 185 124 TVVGTFILPYNYSVFLFSLFTLGIYGLISTFSMSLFGSFQTLKKHQDLPMIIMAVL 183
Qy	186 LVYCAVRI 193
Dd	184 LVFSAVSI 191

	Query Match	Similarity	31.2%	Score	310.5	DB 2.0	Length	194;	
	Best Local	Similarity	36.4%	Pred.	No. 1.5e-17				
	Matches	Conservative	39;	Mismatches	79;	Indels	1;	Gaps	1.
OY	8	AFTWTLTMTAPGENTNIIATSSATSHSGFROSTRVLACMSGFLIWMLCAGISFSIAVI	67						
Db	5	SFLPLVILICSPFGNNFIAMTYANQHGLKSMQCFGVAFQGFLLTSLSFFNIIVLINI	64						
OY	68	DPAAHILSMGAATIVWLAKTIASPRKEQLQAKPIS - FASALIQFNWKILLYVT	126						
Db	65	LPIIEPKLIIGVAAMLTYAEFKILTSKTSADDERHNKNLFVVGIFLPVNMPKGILLFGLT	124						
OY	127	ALSTFVLFOQTALSWVWGVSYLAMIIGTFGNVCMAIAGHLFORLEROYGRQINILALLL	186						
Db	125	VVSTFILPYNYSYSSFLLFSFLTGIVVGLMSTPSWELFCPSMFQKPLLKHITTSNNIMAVULL	184						
OY	187	VYCAYRI	193						
Db	185	VFSATSI	191						

RESULT 13		
Q65N28	PRELIMINARY;	PRT; 192 AA.
ID Q65N28		
AC Q65N28;		
DT 25-OCT-2004 (TREMBLrel. 28, Created)		
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)		
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)		
DE lyse (lysine exporter protein) (LYSE/YGGA) .		
GN Name=LYSE; ORFNames=BL02014, BL100587;		
OS Bacillus licheniformis DSM 13.		
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
OX NCBI_TaxID=279010;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=DSM 13;		
RX PubMed=15383718;		
RA Veith B., Herzberg C., Steckel S., Feesche J., Maurer K.H.,		
RA Ehrenreich P., Baumer S., Henne A., Liesegang H., Weigl R.,		

RESULT 12		
OBID17		
ID	OBID17	PRELIMINARY; PRT; 194 AA.
AC	OBID17:	
DT	01-UN-2003 (TREMBlrel. 24, Created)	
DT	01-UN-2003 (TREMBlrel. 24, Last sequence update)	
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)	
DE	Transporter, Lyse family.	
GN	OrderedlocusNames=BC378:	
OS	Bacillus cereus (strain ATCC 14579 / DSM 31).	
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.	
OX	NCBI_TaxId=225900;	
RN	[1]	
RN	SEQUENCE FROM N.A.	
RX	MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;	
RA	Ivanova N., Sorokin A., Anderson I., Galleron B., Candelon B.,	
RA	Kapitel' V., Bhattacharya A., Resnik G., Mikhailova N., Lapidus A.,	
RA	Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,	

RT Organism with Great Industrial Potential.";
 RL J. Mol. Microbiol. Biotechnol. 7:204-211(2004).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 14580;
 RA Ray M.W., Ramaiya P., Nelson B.A., Brody-Karpin S.D., Zaretsky E.J.,
 RA Tang M., de Leon A.L., Xiang H., Guest V., Clausen I.G., Olsen P.B.,
 RA Rasmussen M.D., Andersen J.T., Jorgensen P.L., Larsen T.S.,
 RA Sorokin A., Bolotin A., Lapidus A., Galleron N., Ehrlich S.D.,
 RA Berta R.M.;
 RT "Complete genome sequence of the industrial bacterium *Bacillus*
 RT *licheniformis* and comparisons with closely related *Bacillus* species.";
 RL Genome Biol. 5:R77(2004).
 DR EMBL, AB017333; AAU9536.1; -
 DR EMBL, CP000002; AAU22181.1; -
 SQ SEQUENCE 192 AA; 21542 MW; 43BDD25276C074EE CXC64;
 Query Match 29.4%; Score 293; DB 2; Length 192;

Best Local Similarity 34.0%; Pred. No. 3.9e-16;
Matches 64; Conservative 43; Mismatches 79; Indels 2; Gaps 2;

QY 8 AFPTTYLTITAMTPGPNNTLLSSATSHGFROSTRVLAMGSLGFLIWLACAGISFSLAVI 67
DB 5 AFISYVMTSTITPGSPNIMMNNENQRFSGFSGWRFSSGIIAGFAVLGILGCAFTISLYNW 64
QY 68 DPAAVHLISWAGAAYIWLAMKIA-TSPTEKEDGLQAKPISFWMASFALOFVNVKIIIGVT 126
DB 65 IPVEPEFKLAGACYLILYLALQVGTAKKQDSTREAS-SFISGFIQILNINISILFPLT 123
QY 127 ALSTFVLPTQALSMWVGVSVLAMIGTFGNVCWALAGHLFQRLFRQYGRQNLIVALL 186
DB 124 VMSAFVLPFNHSLKSTIYVLAIVLGLMLLLMSGFGSVFKRVFAKHDRVFRVLMCLL 183
QY 187 VYCAVRIF 194
DB 184 MYSASVIF 191

RESULT 14

082971

ID 082971 PRELIMINARY; PRT; 198 AA.

AC 082971;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE A2-5a orf2; hypothetical protein.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20261044; PubMed=10803899;
RX Ondan K., Kuriki T., Takata H., Okada S.;
RT "Cloning of the cyclodextrin glucanotransferase gene from alkalophilic
RT Bacillus sp. A2-5a and analysis of the raw starch-binding domain.";
RL Appl. Microbiol. Biotechnol. 53:430-434(2000).
DR EMBL: AB015670; BAA31530.1; -
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005293; F:Lysine permease activity; IEA.
DR GO: GO:0006865; P:amino acid transport; IEA.
DR InterPro: IPR001123; LysE.
DR Pfam: PF01810; LysE; 1.
KW Hypothetical protein.
SQ SEQUENCE 198 AA; 22112 MW; FE2735005FE97D39 CRC64;

Query Match 27.9%; Score 277.5; DB 2; Length 198;

Best Local Similarity 30.1%; Pred. No. 7.3e-15;

Matches 58; Conservative 46; Mismatches 80; Indels 9; Gaps 2;

QY 8 AFPTTYLTITAMTPGPNNTLLSSATSHGFROSTRVLAMGSLGFLIWLACAGISFSLAVI 67
DB 5 SFLLFVFNSTFTGPNFMAMSPGNKYLKFKICGCVGVFVALLLCSFFVVLINF 64
QY 68 DPAAVHLISWAGAAYIWLAMKIA-TSPTEKEDGLQAKPIS-----FMASFALOFVNVKI 120
DB 65 LPTEIEPLTILGVGYMLYLAFTLTS--KODNNVNDGNDGSDNNLFTVMGVLLQFLNPKG 122
QY 121 ILVGVTAISTFVLPTQALSMWVGVSVLAMIGTFGNVCWALAGHLFQRLFRQYGRQNL 180
DB 123 VLVGITVATFLFPYVNSYFSLIFSLFLGVGLMSSCSWSLFGSIFKKFLVQYRKPFNI 182
QY 181 VVALLVYCAVRI 193
DB 183 IMAVLLIYSAISI 195

RESULT 15

0911G9

ID 0911G9 PRELIMINARY; PRT; 205 AA.

AC 0911G9;
DT 01-MAR-2001 (TReMBLrel. 16, Created)

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)

DE Hypothetical protein.
GN OterredlocusNames=PA2306;
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043; DOI=10.1038/35023079;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino R., Wesbrock-Wadman S., Yuen Y.,
RA Brody L.B., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RT Nature 406:959-964(2000).
DR EMBL: AE004657; MAG05694.1; -
DR PIR: B83356; B83356.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005293; F:Lysine permease activity; IEA.
DR GO: GO:0006865; P:amino acid transport; IEA.
DR InterPro: IPR001123; LysE.
DR Pfam: PF01810; LysE; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 205 AA; 22082 MW; 0363C09C75777996 CRC64;

Query Match 27.4%; Score 272.5; DB 2; Length 205;

Best Local Similarity 37.0%; Pred. No. 1.9e-14;

Matches 71; Conservative 39; Mismatches 67; Indels 15; Gaps 8;

QY 5 LLSAFVYTLITAMTPGPNNTLLSSATSHGFROSTRVLAMGSLGFLIWLACAGISFSL 64
DB 7 LLAFSLFPAFVISTFTGPNNTMLASGVNGFVRSIIPHILIGISGFFIWM--AVGFGL 63
QY 65 AVI--DPAAVHLISWAGAAYIWLAMKIA-TSPTEKEDGLQAKPISFWMASFALOFVNVK 119
DB 64 GTVEFAYPVLYTLIRVGAAYIYLWLAMKIA-TSPTEKEDGLQAKPISFWMASFALOFVNVK 123
QY 120 ILVGVTAISTFVLPTQALSMWVGVSVLAMIG--TFGNVCWALAGHLFQRLFR--QYG 175
DB 124 AMWAVGAISTYF-PLQGFITNVVISTVFALINAPTI--IWAGFGSMILRNVLDRPRL 180
QY 176 RQNLIVALLLV 187
DB 181 RVFNVGMALLLV 192

Search completed: September 10, 2005, 01:45:10

Job time: 177 secs

This Page Blank (uspto)